

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieure et de la Recherche Scientifique  
UNIVERSITY ABBÈS LAGHROUR KHENCHELA  
FACULTE DES SCIENCES ET TECHNOLOGIE  
DEPARTEMENT D'INFORMATIQUE

**Mémoire de fin d'étude pour l'obtention du diplôme  
Mastre en Informatique  
Option : Securète et Technologie de Web**

**Thème :**

**Apprentissage automatique pour  
l'évaluation des qualités d'images**



**Réalisé par :**

- BERKANI CHAABANE
- BAKHA HICHEM

**Encadré par :**

- Dr. AHMED SEGHIR ZIANOU

**Le jury:**

- Dr. ABBAS FAYÇAL
- Dr. CHERGHI OUTAILA

*Promotion : 2019/2020*



# *Remerciement*

- *Nous remercions tout d'abord Allah qui nous éclaire le bon chemin*
- *Nous tenons à remercier infiniment Docteur AHMED SEGHIR ZIANOU pour Nous avoir encadrés durant la réalisation de ce travail, Sa grande connaissance du domaine, ainsi que son expérience, ont joué un rôle important pour l'avancement et l'aboutissement de projet, Veuillez trouver ici l'expression de ma respectueuse considération et ma profonde admiration pour toutes vos qualités scientifique et humain*
- *On remercie vivement les docteurs Abbas Fayçal et Cherghi Outaila pour avoir accepté de juger ce mémoire, qu'ils trouvent ici l'expression de nos profondes reconnaissances.*
- *Et tous nos respects et nos profondes reconnaissances à tous ceux qui ont aidés et encouragés à réaliser ce travail.*

# *Dédicace*

*Je dédie ce modeste travail à tous ceux qui sont sacrifiés*

*Pour que notre vie soit meilleure.*

*A mes chers parents(laalia, abdelazize).*

*Mes très chers enfants (nourhane, layane) et leur mère.*

*A mes très chers frères et mes sœurs et je leurs souhaite une bonne  
vie dans tous les domaines.*

*A mon amis baazize lebouzide.*

*Tous mes amis , sans exception.*

*A mon collègue : HICHEM BAKHA.*

*A toute la promotion 2019/2020 .*

*CHAABANE(sofiane)*

# ***Dédicace***

*Je dédie ce modeste travail à :*

*Mes très chers parents qui m'ont encouragé tout le long de mon cycle d'étude et m'ont guidé vers la réussite.*

*Mes très chers enfants et leur mère.*

*Mes frères, et mes sœurs et je leurs souhaite la réussite dans leur vie et le bonheur pour leurs enfants.*

*Tous mes amis, sans exception.*

*Tout personne qui me connais*

*A mon collègue : BERKANI CHAABANE*

*Tous mes collègues de la promotion 2019/2020.*

*Tout personne qui nous a aidés a réalisé ce travaille.*

*Toi qui parcourir ce travaille.*

*HICHEM BAKHA*

## TABLE DES MATIERES

Liste des tableaux

Table des figures

Liste des abréviations

Introduction générale.....1

### Chapitre 1 : Généralités sur le traitement d'image

1.1 Introduction .....	4
1.2. Qu'est qu'une image?.....	4
1.3 Caractéristique d'une image : .....	4
1.4 Types d'images : .....	6
1.4.1 Images Noir et Blanc (monochromes) : .....	6
1.4.2 Images couleur : .....	7
1.5 Système de traitement d'images : .....	8
1.6 Extraction de contours : .....	10
1.7 Segmentation en régions : .....	11
1.7.1 Accroissement de région : .....	11
1.7.2 Méthodes du type "split and merge" : .....	12
1.8 Approches coopératives en traitement d'images : .....	14
1.9 Domaine d'application du traitement d'image : .....	15
Conclusion.....	16

### Chapitre 2 : Evaluation de la qualité de l'image

2.1 introduction.....	18
2.2 l'évaluation de la qualité d'image.....	19
2.2.1 introduction.....	19
2.3 la fidélité.....	20
2.4 Différents types de dégradations .....	20
2.4.1 Dégradation de flou .....	20
2.4.2 Effets de bloc .....	23
2.4.3 Contraste .....	23
2.5 Évaluation subjective.....	24
2.5.1 Protocoles d'évaluation subjective .....	25
2.5.1.1 Simple stimulus .....	25
2.5.1.2 Double stimuli : .....	26
2.5.1.3 Méthode Comparatif .....	27
2.5.2 Bases d'images dédiées à l'évaluation de performance... ..	27
2.6 Évaluation objective .....	31

### Chapitre 3 : les images et vidéos stéréoscopiques

3.1-Introduction .....	34
3.2-La vision humaine tridimensionnelle.....	34
3.3 - la Stéréoscopie .....	36

3.3.1- la vue en relief .....	36
3.4 - c'est quoi la stéréoscopie.....	36
3.5 l'Histoire de la Stéréoscopie .....	38
3.6- les différents types de stéréoscope.....	41
3.6.1 stéréoscope de miroirs .....	41
3.6.2 stéréoscope à lentilles ou à prismes .....	42
3.6.3 stéréoscope à prismes .....	43
3.6.4 l'ère du numérique .....	43
3.7 Comment faire de la Stéréoscopie ?.....	45
3.8 Les images stéréoscopiques.....	46
3.9 Comment créer une image 3D stéréoscopique ?.. .....	46

**Chapitre 4 : Évaluation de la qualité d'image et vidéo stéréoscopique sans référence basée sur les entropies spatiales et spectrales de l'image**

Introduction .....	49
4.1 machines à vecteurs de support (SVM) .....	50
4.1.1 Introduction .....	50
4.1.2 Notions sur l'apprentissage.....	50
4.1.3 Principe des machines à vecteurs de support .....	51
A) Cas de données linéairement séparables.....	51
B) Cas de données non linéairement séparables .....	58
4.1.4 SVM pour la régression(SVR) .....	60
4.2 Différents modèles de combinaison binoculaire.....	61
4.3 Carte de disparité.....	63
4.3.1 définition de disparité .....	63
4.4 Modèle d'activité spatiale binoculaire .....	64
4.4.1 Fusion binoculaire, rivalité et suppression .....	65
4.4.2 Information spatiale et perception binoculaire .....	65
4.4.3 Modèle d'activité spatiale binoculaire .....	67
4.4.4 évaluation de la qualité des stereopaire sans référence.....	68
4.5 Caractéristiques d'entropie d'image.....	69
4.5.1 Caractéristiques de l'entropie spatiale .....	70
4.5.2 Caractéristiques de l'entropie spectrale .....	71
Conclusion.....	73

## **Chapitre 5 : Résultats et Discussions**

5.1 Introduction : .....	75
5.2 Base de qualité de vidéo 3D Waterloo-IVC Phase I: .....	75
5.3 Environnement de programmation et d'implémentation.....	76
5.4 Critère d'évaluation de qualité d'image (vidéo).....	77
5.4.1 Indice d'ajustement de qualité.....	77
5.4.2 Coefficient de corrélation.....	77
5.4.3 Coefficient de corrélation des rangs de Spearman (ROCC)... ..	78
5.4.4 Erreur quadratique moyenne (RMSE).....	78
5.4.5 Significations des Coefficients.....	79
5.5 Résultat de méthode proposée.....	79
5.5.1 Performance et comparaison.....	79
5.6 conclusion.....	81
<b>Conclusion générale.....</b>	<b>83</b>
Bibliographie.....	84

---

<b>TABLEAU</b>	<b>PAGE</b>
2.1 – Echelle d'évaluation de qualité d'image.....	26
2.2 – Echelle d'évaluation de qualité d'image à 5 niveaux.....	26
2.3 – Echelle d'évaluation de la qualité.....	27
4.1: caractéristiques utilisées pour SSEQ (dubbed Spatial–SpectralEntropy-based Quality).....	73
5.1 La base WATERLOO-IVC 3D VIDEO PHASE I.....	75
5.2: Comparaison des performances des modèles de prédiction de qualité 2d à 3d sur la base WATERLOO-IVC 3D VIDEO.....	81

<b>FIGURES</b>	<b>PAGE</b>
1.1 image de pixel.....	5
1.2 Exemple de texture.....	5
1.3 Histogramme et palette associés à une image.....	6
1.4 image binaire.....	7
1.5 image en niveaux de gris.....	7
1.6 image couleur.....	8
1.7 Schéma d'un système de traitement d'images.....	8
1.8 Egalisation d'histogramme.....	9
1.9 Différents types de contours : marche, toit et pointe.....	10
1.10 Voisinage d'un pixel.....	12
1.11 Découpage par quadtree d'une image 8x8. Traitement avec seuil=100%.....	13
1.12 Problème de découpage arbitraire de régions dans la phase de "split".....	13
1.13 Principes de la croissance de région et du suivi de contour. Pour faire croître la primitive (pixels foncés), différents candidats sont possibles (pixels clairs). Le pixel ayant la meilleure évaluation sera ajouté à la primitive.....	15
1.14 Boucle de contrôle du comportement de perception. Ce comportement boucle sur Une liste de pixels tant que celle-ci n'est pas vide.....	15
2.1 Représentation des différentes phases de traitement d'une image.....	18
2.2 L'effet de dégradation, (a)Originale, (b)Dégradée.....	19
2.3 Flou de défocalisation.....	20
2.4 Exemple de flou de mouvement.....	21
2.5 Flou dû à une perte d'information.....	21
2.6 Flou dû à la compression, (a)Image compressée par le JPEG2000, (b)Image compressée par le JPEG.....	22
2.7 Un flou dû à un filtrage.....	22
2.8 Flou dû à la transmission.....	22
2.9 Images compressées par JPEG : (a) Effets de blocs horizontal, (b) Effets de blocs vertical.....	23
2.10 Images représentant la dégradation contraste.....	24
2.11 Méthode à simple stimulus.....	25
2.12 méthode a double stimulus.....	26
2.13 Une échelle comparative de qualité visuelle .....	27
2.14 Images de référence de la base LIVE.....	28
2.15 Images de référence de la base CSIQ.....	29
2.16 Images de référence de la base IVC.....	29
2.17 Images de référence de la base TID.....	30
2.18 Images de référence de la base LIVE 3D phase I.....	31

2.19 Exemple de relation entre les scores objectifs et subjectifs.....	32
3.1 L'œil humain.....	34
3.2 capture de l'image pas les deux yeux.....	35
3.3 deux images capturée par les deux yeux.....	35
3.4 une vision 3D de l'humain.....	36
3.5 Ces deux images sont transférées au cerveau, qui en « fabrique » une troisième en relief (3D).....	37
3.6 Principe de la vision stéréoscopie.....	37
3.7 Stéréoscope de Wheatstone.....	38
3.8 Stéréoscope de Jules Duboscq.....	38
3.9 Verascope.....	39
3.10 Appareil à film 24x36, Verascope 40 de Jules Richard.....	39
3.11 Photo prise sur Mars.....	40
3.12 Photo du soleil.....	40
3.13 Avatar.....	40
3.14 Caméra 3D Professionnelle.....	41
3.15 Stéréoscope de Wheatstone.....	41
3.16 Cartographie.....	42
3.17 stéréoscope à lentilles.....	42
3.18 stéréoscope de Holmes.....	43
3.19 Pokescope.....	43
3.20 Lunettes actives Nvidia avec écran Samsung.....	44
3.21 Lunettes actives (ou à LCD).....	44
3.22 Lunettes passives (ou polarisantes).....	45
3.23 Canon montés ensemble et synchronisés par le logiciel StereoData Maker .....	46
3.24 constitution d'images stéréoscopiques 3D.....	47
4.1 a) Cas linéairement séparable, b) Cas non linéairement séparable.....	51
4.2 Exemple d'un hyperplan séparateur.....	52
4.3 Exemple de vecteurs de support.....	52
4.4 Exemple de recherche d'un hyperplan optimal.....	53
4.5 Exemple de marge maximale (hyperplan valide).....	53
4.6 SVM binaire.....	53
4.7 a) Hyperplan avec faible marge, b) Meilleur hyperplan séparateur.....	54
4.8 Exemple de classification d'un nouvel élément.....	54
4.9 Exemple de changement de l'espace de données.....	58
4.10 Illustration de cas non linéairement séparable (le cas XOR).....	59
4.11 Illustration de passage d'un espace 2D à un espace 3D.....	59

---

4.12 Hyper-tube modélisant la fonction de régression.....	60
4.22 l'image de disparité.....	64
4.23 Nuage de points de DMOS par rapport aux scores MS-SSIM correspondants des vues gauche / droite.....	66
4.24 Nuage de points de DMOS par rapport aux scores MS-SSIM correspondants des vues gauche / droite.....	67
4.25 diagramme du modèle proposé d'évaluation de la qualité NR(sans référence) des images S3D.....	69
4.26 Histogrammes des valeurs d'entropie spatiale pour différents types de Distorsion.....	72
4.27 Histogrammes des valeurs d'entropie spectrale pour différents types de Distorsion.....	73
5.1 Exemples d'images de vidéos. Seules les vues de droite sont affichées.....	76
5.2 Evaluation de la qualité.....	78
5.3 Plot des scores objectifs pour les images avec les MOS correspondant.....	80

CCD : Dispositif de transfert de charge (charge-coupled device)

SVH : Système visuel humain

MOS : Score opinion moyen

DMOS: (Différence) Score opinion moyen

3D: 3 Dimension

EQI : L'évaluation de la qualité de l'image

MAR-EQI : Métriques avec référence de L'évaluation de la qualité de l'image

(MRR-EQI). Métriques avec référence réduite de L'évaluation de la qualité de l'image

(MSR-EQI). Métriques sans référence réduite de L'évaluation de la qualité de l'image

S3D : Stéréoscopique 3D

(SVM) : Machine à vecteur de support

(SVR) : Régression à vecteur de support ,séparateur à vaste marge

SSIM : L'indice de similarité structurel

DCT : Transformée en cosinus discrète

IQA : Evaluation de la qualité d'image

SSEQ (Qualité basé sur entropie spatial et spectrale)

MEQ : Métriques d'évaluation de qualité

NSS : Les statistiques de scènes naturelle

ML: Machine learning

SROCC: Spearman linear correlation coefficient

RMSE: La moyenne quadratique de l'erreur

JP2K : Distorsion de la compression JPEG2000

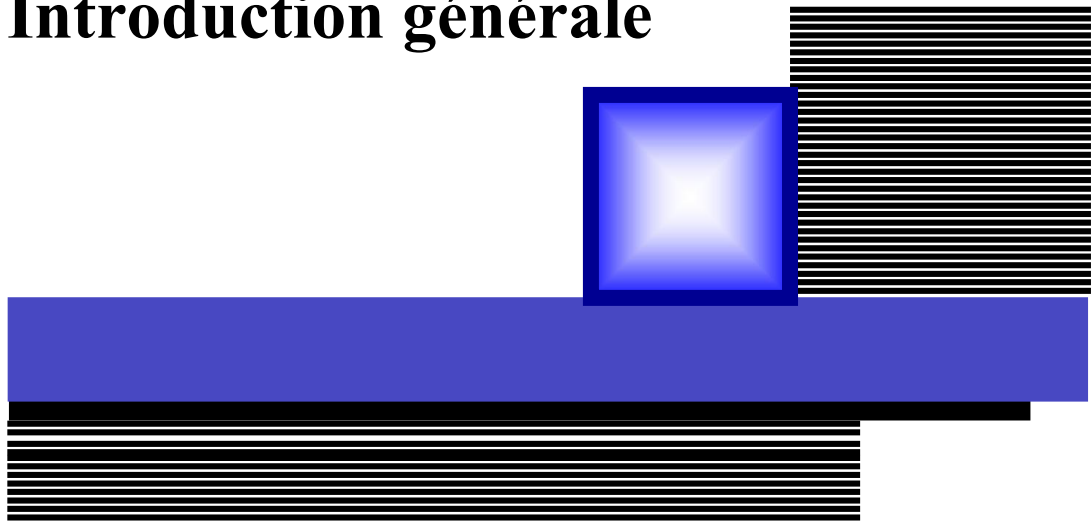
JPEG : Distorsion de la compression JPEG

FF : Bruit du à la transmission (faste fading)

CC : Coefficient de corrélation Pearson

ROCC : Coefficient de corrélation des rangs de Spearman

# Introduction générale



## **Introduction générale :**

Les images ne jouent pas seulement un rôle d'accompagnement dans notre vie : elles attirent l'attention de l'œil et facilitent la compréhension, comme dit le proverbe une image vaut mille mots. L'image représente également le contenu le plus répandu et le plus utilisé, on la trouve dans les téléphones, la télévision et sur les réseaux sociaux. En effet, l'être humain a tendance à enregistrer, répertorier ou à partager les moments qu'il aime ou qui le touche. Il est donc important de choisir des images avec une meilleure qualité.

Grâce aux avancées technologiques, l'image a pris une place prépondérante dans le quotidien des sociétés contemporaines. En effet, des équipements sophistiqués ont été développés pour l'acquisition et la diffusion. La haute résolution de ces équipements se traduit par des tailles importantes des scènes capturées. Ceci nécessite des espaces de stockage importants et fait explicitement appel aux différentes méthodes de compression.

Ce développement a engendré à la fois une amélioration de la qualité, et également un gain de temps précieux. En revanche, il est aussi la source de nouvelles problématiques qui n'existaient pas auparavant.

Tout d'abord, les systèmes d'acquisition génèrent des dégradations telles que le bruit et le flou. D'autre part, il y a les artéfacts dus à la compression et à la transmission.

Prenant en compte la sensibilité du système visuel humain, qui est très exigeant en terme de qualité, il est impératif d'avoir des systèmes d'évaluation capables de mesurer les effets de la dégradation sur l'image. Le premier recours s'oriente vers les mesures subjectives où on fait appel à un groupe d'observateur qui évalue la qualité des images. Ce type d'évaluation n'est pas pratique, étant donné qu'il suggère davantage en terme de ressources et du temps.

Évaluer la qualité des images sans faire appel aux observateurs est une solution envisageable. On nomme ce type d'évaluation, les métriques objectives. Elles cherchent de remplacer l'être humain tout en maintenant des capacités d'évaluation similaires.

Pour cela, des algorithmes utilisant l'information contenue dans le signal image ont été mises en œuvre. La dernière décennie a connu un grand intérêt pour la qualité objective avec des dizaines de métriques qui ont été proposées. Les mesures objectives de qualité d'image peuvent être classées en trois groupes : les mesures de qualité avec référence, les mesures de qualité avec référence réduite et les mesures de qualité sans référence.

Le classement est basé sur la présence de l'image de référence. Pour les deux premiers groupes, la totalité ou une partie de l'image est présente tandis que pour le dernier groupe, l'image de référence est totalement indisponible.

Dans le cadre de ce mémoire, nous nous sommes intéressés à une métrique objective générique de qualité sans référence sur une base de vidéo stéréoscopique 3D, cette métrique est basé sur les caractéristique d'entropie spatiale et spectrale de l'image dégradé par des différents types distorsions

affectant sa qualité, avec l'incorporation de la vision binoculaire, où l'image est perçue à partir de la combinaison de la vue gauche et la vue droite d'une image.

Une machine d'apprentissage SVM est utilisée pour prédire ou dériver un score de qualité.

Le mémoire est composé de 5 chapitres :

**Le chapitre 1 est une généralité sur les traitements d'images :** Cette partie représente des notions de base sur le traitement d'image, Une définition de l'image est donnée au début, après les caractéristiques de l'image sont abordées ainsi que les types d'images sont présentées, Ensuite les différents prétraitements effectués sur l'image sont présentés aussi, et enfin les domaines d'application du traitement d'image.

**Le chapitre 2 est une introduction à la qualité d'image :** Cette partie représente une introduction à l'évaluation de la qualité d'image. Une définition est donnée au début. Après, les différentes dégradations seront abordées, ainsi que les artefacts qui sont introduits par chacune. Ensuite, l'évaluation subjective et objective sont présentées.

**Le chapitre 3 concerne les images et vidéos stéréoscopiques :** Dans ce chapitre on va voir les images et vidéos 3D, leur différence avec 2D. en premier lieu La vision humaine tridimensionnelle est présentée ainsi que les concepts de stéréoscopie et leur histoire sont abordés, ensuite une définition de l'image stéréoscopique est présentée, et aussi comment faire de la stéréoscopie.

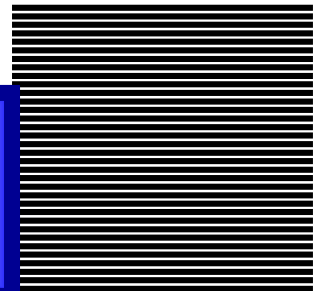
**Le chapitre 4 est dédié à la métrique proposée :** Dans ce chapitre, une nouvelle métrique sans référence est proposée qui est basée sur l'entropie spatiale et spectrale de l'image avec l'incorporation de la vision humaine binoculaire sur la base d'images (vidéos) stéréoscopiques 3D, Premièrement la théorie des SVR est détaillée, Ensuite Différents modèles de combinaison binoculaire ainsi que les Cartes de disparité sont présentés, et enfin les caractéristiques d'entropie spatiale et spectrale de l'image sont abordées.

**Le chapitre 5 présente les résultats et discussions de l'application de la métrique :** la métrique est implémentée par un outil de développement Matlab, la bibliothèque SVM LIBSVM est utilisée pour le SVR, dans ce chapitre une évaluation de la performance de la métrique est présentée ainsi qu'une étude comparative avec d'autres métriques sur une base de vidéos 3D Waterloo-IVC Phase I est présentée. La métrique donne des résultats significatifs et prometteurs.

# Chapitre

*Généralités sur le traitement  
d'images*

1



# Chapitre 01 : Généralités sur le traitement d'images

## 1.1 Introduction

Le traitement d'images est un domaine très vaste qui a connu, et qui connaît encore, un développement important depuis quelques dizaines d'années.

On désigne par traitement d'images numériques l'ensemble des techniques permettant de modifier une image numérique afin d'améliorer ou d'en extraire des informations.

De ce fait, le traitement d'images est l'ensemble des méthodes et techniques opérant sur celles-ci, dans le but de rendre cette opération possible, plus simple, plus efficace et plus agréable, d'améliorer l'aspect visuel de l'image et d'en extraire des informations jugées pertinentes.

Dans ce chapitre, nous abordons les notions de base nécessaires à la compréhension des techniques de traitement d'images. Ensuite, nous allons donner un aperçu sur les différents techniques connues dans ce domaine.

## 1.2 -Qu'est qu'une image ?

Une image est un ensemble de points appelés pixels, ce pixel est contenu dans un tableau à deux dimensions donnant l'intensité lumineuse en chaque pixel de coordonnées spatiales  $(x,y)$ .

L'image est une représentation d'une personne ou d'un objet par la peinture, la sculpture, le dessin, la photographie, le film, etc.

C'est aussi un ensemble structuré d'information qui, après affichage sur l'écran, a une signification pour l'œil humain.

Elle peut être décrite sous la forme d'une fonction  $I(x,y)$  de brillance analogique continue, définie dans un domaine borné, tel que  $x$  et  $y$  sont les coordonnées spatiales d'un point de l'image et  $I$  est une fonction d'intensité lumineuse et de couleur. Sous cet aspect, l'image est inexploitable par la machine, ce qui nécessite sa numérisation.

## 1.3 Caractéristique d'une image :

L'image est caractérisée par un ensemble de paramètres parmi lesquels on peut citer les suivants : pixel, intensité, bruit, résolution, contour, texture, histogramme, Contraste.

**1.3.1 le pixel :** Le pixel est le plus petit point de l'image si le bit est la plus petite unité d'information que peut traiter un ordinateur. Dans une image couleur (R.V.B), un pixel peut être représenté sur trois octets : un octet pour chacune des couleurs : rouge (R), vert (V) et bleu (B).



Figure 1.1 : image de pixel.

**1.3.2 Intensité :** L'intensité ou luminance est le caractère qui indique l'intensité de lumière perçues indépendamment de la couleur. Elle s'étend du noir au blanc avec toutes les nuances de gris si on ne voit pas la couleur.

**1.3.3 bruit :** Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur.

**1.3.4 Résolution :** La résolution d'une image numérique est définie par un nombre de pixel par unité de longueur de la structure à numériser. Ces paramètres sont définis lors de la numérisation et dépend principalement des caractéristiques du matériel utilisé lors de processus de numérisation. Plus le nombre de pixel est élevés par unité de longueur de la structure à numériser, plus la quantité d'information qui décrit cette structure est important et plus la résolution est élevée.

**1.3.5 Contour :** Un contour est défini comme une frontière entre deux régions différentes (suivant un critère similarité donné, par exemples un niveau de gris identique à une couleur identique ...).

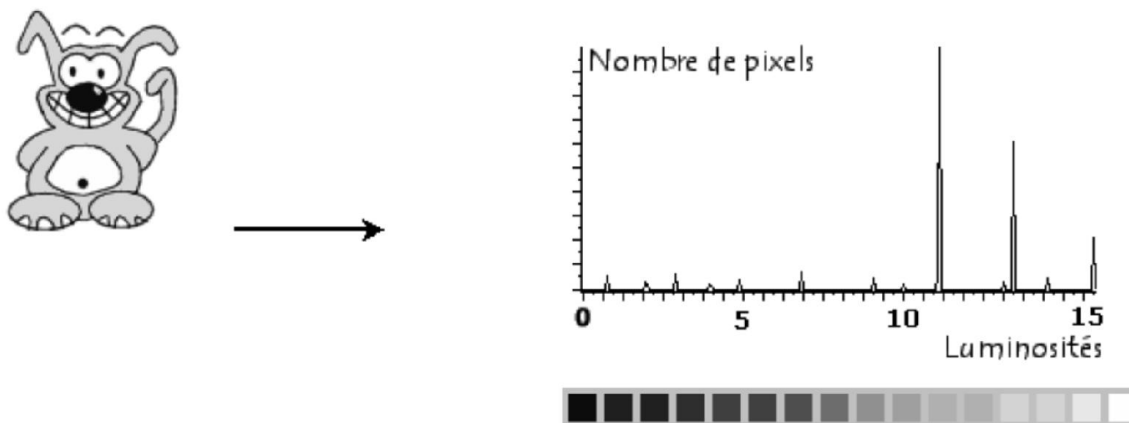
**1.3.6 Texture :** La définition littéraire de la texture est la suivante : " répétition spatiale d'un même motif dans différentes directions de l'espace ". Cette définition est limitative car elle caractérise l'objet indépendamment d'une observation humaine. La notion de texture est utilisée pour traduire un aspect homogène de la surface d'un objet sur une image.



Figure 1.2 : Exemple de texture.

**1.3.7 histogramme :** L'histogramme des niveaux de gris ou des couleurs d'une image est une fonction qui donne la fréquence d'apparition de chaque niveau de gris (couleur) dans l'image. Il permet de donner un grand nombre d'information sur la distribution des niveaux de gris (couleur) et de voir entre quelles bornes est répartie la majorité des niveaux de gris (couleur) dans le cas d'une image trop claire ou d'une image trop foncée.

Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire les informations utiles de celle-ci. Pour diminuer l'erreur de quantification, pour comparer deux images obtenues sous des éclairages différents, ou encore pour mesurer certaines propriétés sur une image, on modifie souvent l'histogramme correspondant.



**Figure 1.3** Histogramme et palette associés à une image

**1.3.8 Contraste :** C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d'images. Si  $L_1$  et  $L_2$  sont les degrés de luminosité respectivement de deux zones voisines  $A_1$  et  $A_2$  d'une image, le contraste  $C$  est défini par le rapport :  $C = (L_1 - L_2) / (L_1 + L_2)$

## 1.4 Types d'images :

Il ya différents types d'images sur ordinateur on peut distinguer :

### 1.4.1 Images Noir et Blanc (monochromes) :

Ces images sont dites à niveaux de gris, car on ne prend pas en compte la couleur mais seulement l'intensité lumineuse, parmi ces images on peut trouver :

**a-Images binaires :** Une image binaire est une image pour laquelle chaque pixel ne peut avoir valeur que 0/1.



**Figure 1.4 :** image binaire .

**b- Images en niveaux de gris :** Dans ce cas , on dispose d'une échelle de teinte grise, et dans la plupart de temps on dispose 256 niveaux de gris avec : 0 → noir,.....127 → gris moyen,...255 → blanc, Ceci est commode car l'unité d'information est l'octet certaines images peuvent être codées sur deux octets ou plus (exp : images médicales, des images astronomiques...)



**Figure 1.5 :** image en niveaux de gris

### 1.4.2 Images couleur :

Ces images sont en générale codées en utilisant le codage des trois couleurs Fondamentales (rouge, vert, bleu) on parle alors d'images RVB. Chaque couleur est codée sous forme d'un octet d'où :

Image RVB → Composant Rouge, intensité de 0 à 255.  
 Composant Verte, intensité de 0 à 255.  
 Composant Bleue, intensité de 0 à 255.



Figure 1.6 : image couleur.

## 1.5 Système de traitement d'images :

Un système de traitement numérique d'images est composé de :

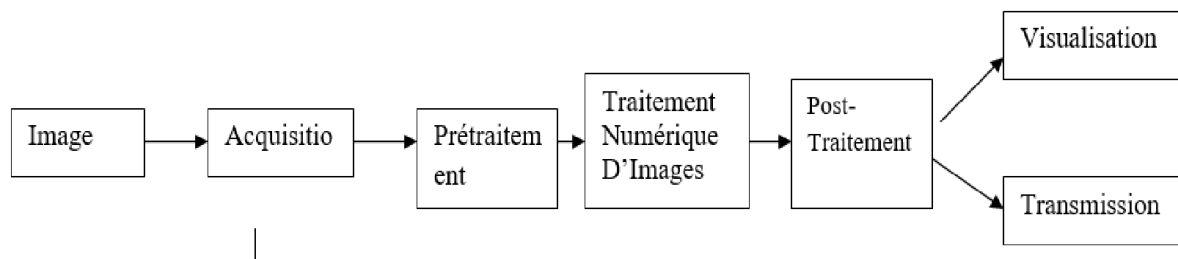


Figure1.7 Schéma d'un système de traitement d'images

### 1.5.1 Prétraitement d'images :

Cette phase a lieu juste après l'acquisition des images et a pour objectif d'améliorer la qualité de l'image en vue de sa segmentation.

Le temps de traitement (temps CPU) est très important. C'est un facteur décisif, il doit être le plus petit possible. Ceci implique que les opérateurs doivent être locaux, c'est-à-dire qu'ils doivent intervenir sur un nombre limité de pixels et essentiellement sur des pixels dans le voisinage proche du pixel courant.

Les techniques de prétraitements les plus courantes qu'on va présenter sont :

- La modification d'histogramme.
- La réduction du bruit par filtrage.

#### 1.5.1.1 Modification d'histogramme :

On cherche à améliorer l'image en appliquant une transformation ponctuelle d'intensité, à tout pixel  $a_s(x,y)$  on associe une intensité par la transformation  $T$  croissante de façon à conserver des contrastes entre régions (claire sur fond sombre, apparaît plus claire après transformation). Le pixel après transformation est le pixel  $a's(x,y)$  :  $a's(x,y) = T(a_s(x,y))$

Dans ce qui suit, on va présenter deux types de modification d'histogramme :

- Expansion de dynamique
- Egalisation d'histogramme

a. **Expansion de dynamique** : Cette méthode consiste à utiliser au mieux l'échelle des niveaux de gris disponible sur le système d'acquisition d'images.

Soit  $A[i,j]$  l'image de départ,  $A'[i,j]$  image après transformation et  $[a_0, a_1]$  l'intervalle des intensités présentes dans l'image et soit  $[a_{\min}, a_{\max}]$  l'intervalle disponible.

L'expansion du dynamique correspond à la transformation  $a_s = \alpha + \beta a_s$

Telle que :

$$\forall a \in [a_0, a_1] \quad a_s \xrightarrow{T} a'_s \in [a_{\min}, a_{\max}]$$

$$\alpha = \frac{a_{\min} \cdot a_1 - a_{\max} \cdot a_0}{a_1 - a_0} \quad , \quad \beta = \frac{a_{\max} - a_{\min}}{a_1 - a_0}$$

b. **Egalisation d'histogramme** : Cette transformation consiste à rendre le plus plat possible l'histogramme de niveaux de gris de l'image.

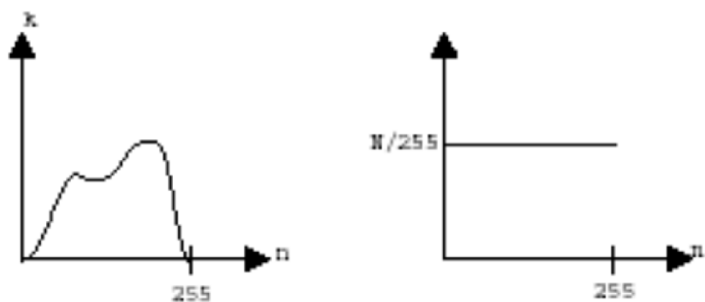


Figure 1.8 Egalisation d'histogramme

### 1.5.1.2 Réduction du bruit par filtrage :

Le bruit peut provenir de différentes causes :

- Environnement lors de l'acquisition,
- Qualité du capteur,
- Qualité de l'échantillonnage.

Plusieurs filtres sont utilisés pour la réduction de bruit, ils sont divisés en deux catégories :

- Filtres linéaires stationnaires,
- Filtres non-linéaires stationnaires.

La première catégorie comprend tous les opérateurs pouvant exprimer leur résultat comme une combinaison linéaire des niveaux de gris d'un voisinage de l'image. Ces filtres possèdent des caractéristiques spectrales, on parle ainsi de filtre passe-bas (l'image devient floue) ou de filtre passe-haut (les contours ressortent).

Voici quelques exemples de filtres réducteurs de bruit :

- Filtre moyenneur
- Filtre gaussien
- Filtre exponentiel de Shen

Le principal inconvénient des filtres linéaire est que la réduction de bruit s'accompagne d'un étalement des transitions entre régions. Ce problème peut être surmonté par l'utilisation des filtres non linéaires.

Le domaine de filtrage non linéaire est extrêmement vaste. Les opérateurs les plus courants en prétraitement d'images sont :

- Les filtres d'ordre,
- Les filtres homomorphique.
- Les filtres morphologiques.

## 1.6 Extraction de contours :

La détection de contour est une étape préliminaire à de nombreuses applications de l'analyse d'images. Les contours constituent en effet des indices riches, au même titre que les points d'intérêts, pour toute interprétation ultérieure de l'image. Les contours dans une image proviennent des :

- Discontinuités de la fonction de réflectance (texture, ombre),
- Discontinuités de profondeur (bords de l'objet),

et sont caractérisés par des discontinuités de la fonction d'intensité dans les images.



**Figure 1.9** Différents types de contours : marche, toit et pointe

Le principe classique de la détection de contours repose sur l'étude des dérivées de la fonction d'intensité dans l'image : les extrema locaux du gradient de la fonction d'intensité et les passages par zéro du laplacien.

## 1.7 Segmentation en régions :

Fondamentalement, la segmentation est un processus qui consiste à découper une image en régions connexes présentant une homogénéité selon un certain critère, comme par exemple la couleur. L'union de ces régions doit redonner l'image initiale, On regroupe généralement les algorithmes de segmentation en trois grandes classes

- Segmentation basée sur les pixels
- Segmentation basée sur les régions
- Segmentation basée sur les contours

La première catégorie travaille sur des histogrammes de l'image. Par seuillage, clustering ou clustering flou, l'algorithme construit des classes de couleurs qui sont ensuite projetées sur l'image. La segmentation est implicite puisqu'on suppose que chaque cluster de l'histogramme correspond à une région dans l'image. En pratique, ce n'est pas forcément le cas et il faut séparer les régions de l'image qui sont disjointes bien qu'appartenant au même cluster de couleur. Ces algorithmes sont assez proches des algorithmes de réduction de couleur.

La deuxième catégorie correspond aux algorithmes d'accroissement ou de découpage de région. L'accroissement de région est une méthode bottom-up : on part d'un ensemble de petites régions uniformes dans l'image (de la taille d'un ou de quelques pixels) et on regroupe les régions adjacentes de même couleur jusqu'à ce qu'aucun regroupement ne soit plus possible.

Le découpage de région est le pendant top-down des méthodes d'accroissement : on part de l'image entière que l'on va subdiviser récursivement en plus petites régions tant que ces régions ne seront pas suffisamment homogènes.

La troisième catégorie s'intéresse aux contours des objets dans l'image. La plupart de ces algorithmes sont locaux, c'est à dire fonctionnent au niveau du pixel.

Les algorithmes que nous allons présenter sont :

- Accroissement de région fonctionnant au niveau du pixel.
- Split and merge classique.

### 1.7.1 Accroissement de région :

Les méthodes d'accroissement de régions sont les méthodes de segmentation les plus simples. Le principe est basé sur une approche bottom-up : l'algorithme part de petits éléments de l'image qu'il va tenter de regrouper en éléments plus importants.

Nous présentons ici la version de base de l'algorithme d'accroissement de région  $\delta$  qui fonctionne en agrégeant des pixels. Supposons une région de couleur homogène  $R$ . Initialement,  $R = 1$  pixel. On va étendre la région  $R$  en incluant les pixels situés sur la frontière et dont la couleur est proche de celle de  $R$  (la variation de couleur est inférieure à un seuil caractéristique de ce type d'algorithmes).

En répétant cette procédure jusqu'à ce qu'il n'y ait plus de pixels de couleur assez proche sur la frontière, on obtient une région de couleur homogène maximale autour du pixel de départ. La région initiale "gonfle" en absorbant des pixels de la frontière, jusqu'à stabilité par rapport à la propriété d'homogénéité. Afin d'assurer un critère de connexité, on utilise les voisins en 4-connexités pour l'accroissement de région et la 8-connexités pour l'accroissement de contour (cas de détection de contours).

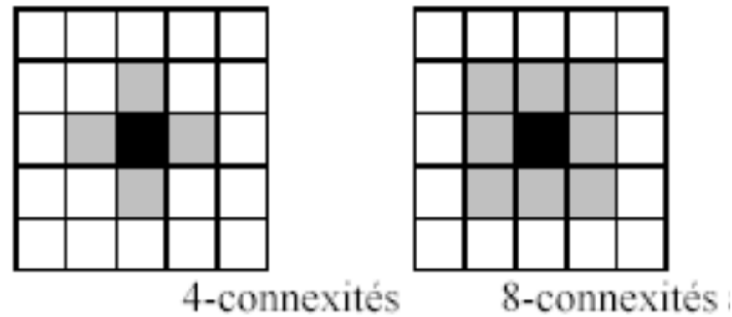


Figure 1.10 Voisinage d'un pixel

Cette méthode présente deux limitations:

- Les régions obtenues dépendent fortement des pixels d'amorçage choisis et de l'ordre dans lequel les pixels de la frontière sont examinés.
- Le résultat final est très sensible à la valeur du  $\delta$  seuil

### 1.7.2 Méthodes du type "split and merge" :

L'algorithme "split and merge" a été présenté la première fois en 1974 par Pavlidis et Horowitz. Cet algorithme s'apparente dans son principe à l'algorithme d'accroissement de région que nous venons de présenter. La différence principale provient de la nature des régions élémentaires agrégées. Dans l'algorithme "split and merge",

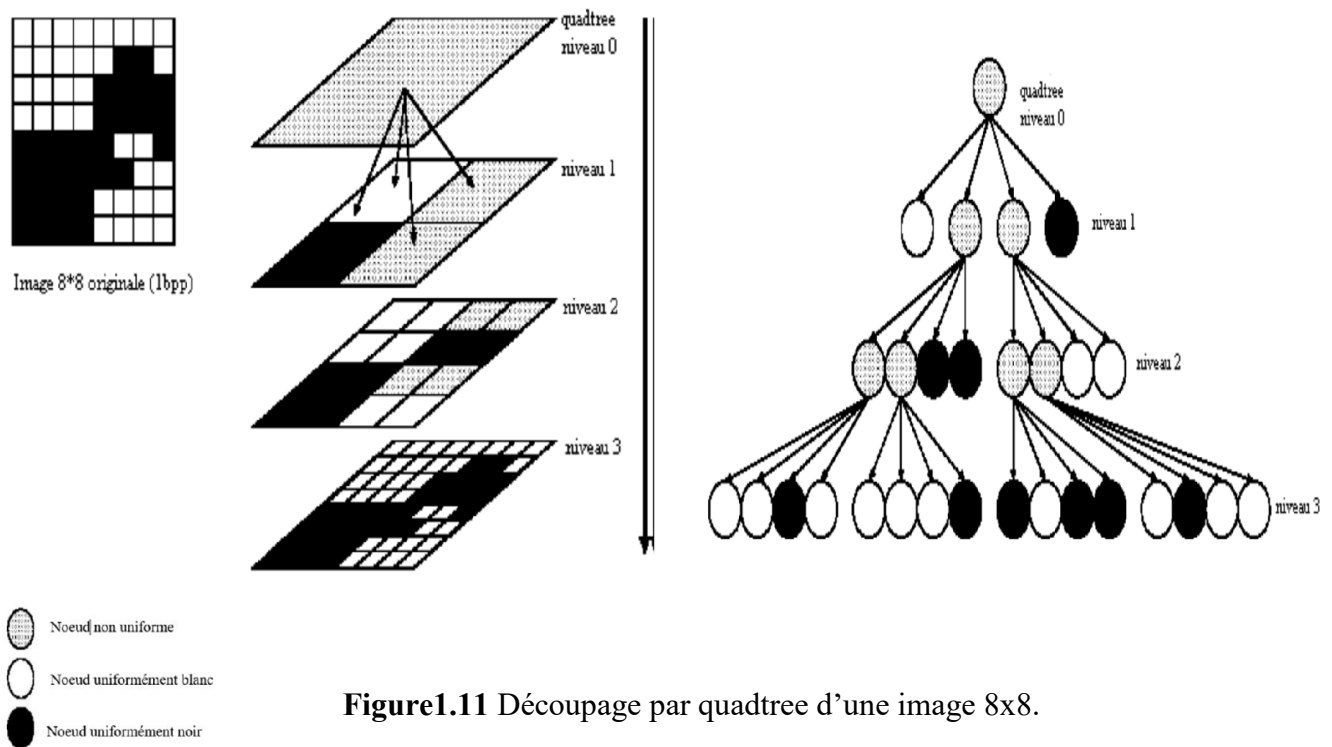
Les régions agrégées proviennent d'une première phase (split) de traitement de l'image qui construit de manière récursive des régions carrées de taille variable mais homogènes.

#### ⇒ Split

La méthode de découpage de l'image utilisée dans cet algorithme est basée sur la notion de "quadtree". Cette structure de données est un arbre quaternaire qui permet de stocker l'image à plusieurs niveaux de résolution. On part d'une région initiale qui est l'image tout entière. Si cette image vérifie un critère d'homogénéité de couleur, l'algorithme s'arrête.

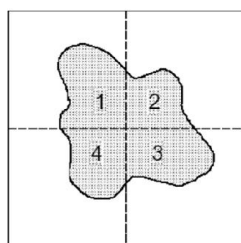
Sinon, on découpe cette région en quatre parties de même taille et on lance la procédure récursivement dans ces quatre parties. La région initiale va être stockée comme un noeud dans un graphe

et les sous parties comme des fils de ce noeud. La figure 1.11 montre une image en noir et blanc 8x8 et le découpage correspondant à chaque niveau, ainsi que la structure d'arbre associée à ce découpage.



### ⇒ Merge

La procédure de découpage décrite précédemment aboutit à un nombre de régions trop élevé. La cause fondamentale de cette sur-segmentation est que l'algorithme découpe les régions de manière arbitraire. Il se peut qu'il coupe de cette façon une zone homogène en deux ou quatre parties (Figure 1.12).



**Figure 1.12** Problème de découpage arbitraire de régions dans la phase de "split"

Les parties 1,2,3 et 4 de la (Figure 1.12) par exemple, appartiennent à des branches différentes du quadtree. Elles sont donc considérées comme des régions différentes bien que leur couleur soit identique.

La solution, qui correspond à la phase “merge” de l’algorithme, est de procéder à une fusion de régions après le découpage. L’implémentation la plus simple de cette fusion cherche tous les couples de régions adjacentes dans l’arbre issu du découpage et cherche à les fusionner si leur couleur est assez proche.

## 1.8 Approches coopératives en traitement d’images :

La résolution collective de problème se préoccupe de développer des systèmes d’agents simples capables de résoudre collectivement un problème complexe ; par exemple construire un chemin reliant tous les noeuds d’un graphe et dont la longueur est minimale (problème dit du voyageur de commerce ou TSP).

La question principale qui se pose alors est de déterminer le système de représentation de l’environnement, les comportements individuels, et la dynamique qui va régir le fonctionnement du système afin qu’il produise une réponse au problème qui lui est posé.

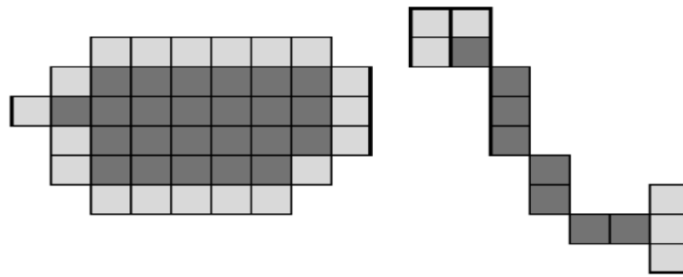
Plusieurs modèles inspirés des insectes sociaux ont déjà été élaborés et utilisés pour effectuer de la résolution de problème. L’une des applications de ces modèles est le traitement d’images.

Dans ce type d’approches, l’image y est vue comme un environnement ou territoire analysé par une société d’agents (ou processus) distribués spatialement. Cette approche est novatrice car elle va à l’encontre de la conception classique centralisée et planifiée des tâches d’analyse d’images.

Ces architectures mettent en relief la distribution de calcul, l’adaptation locale des traitements et la richesse des coopérations possibles entre les aspects contour et région.

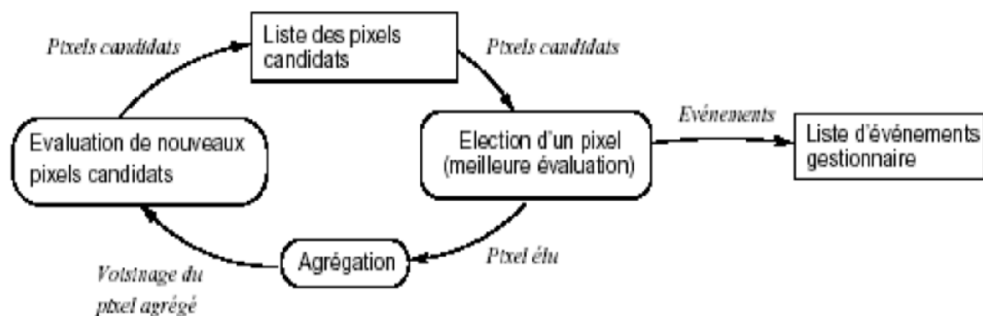
Il existe plusieurs approches coopératives appliquées aux traitements d’images, on peut citer Système de Bellet, Approche de Boucher, Approche de Duchesnay, Approche de Liu & Tang : Segmentation par « agents migrants » .

**1.8.1 Système de Bellet :** Bellet a développé un système coopératif pour la segmentation de bas niveau des images, ce système est multi-processus, puisqu’une image est segmentée par une multitude de processus indépendants, qui sont basés sur des modèles de croissance de région ou de suivi de contour. Chaque processus est initialisé à un endroit précis de l’image, avec un germe, et il fait croître sa primitive à partir de cet endroit. La croissance (région ou contour) s’effectue par une évaluation multi-critères de pixels avoisinants et sélection des meilleurs candidats (figure 1.13).



**Figure 1.13** Principes de la croissance de région et du suivi de contour. Pour faire croître la primitive (pixels foncés), différents candidats sont possibles (pixels clairs). Le pixel ayant la meilleure évaluation sera ajouté à la primitive.

Ce système est basé sur une nouvelle forme de coopération entre différents processus de segmentation de natures différentes. Lorsqu'un processus manque d'informations pour mener à bien les décisions qu'il doit prendre, il crée de nouveaux processus, qualifiés de fils pour segmenter de nouvelles primitives et ainsi, récolter plus d'informations sur l'environnement local. Un processus peut également s'appuyer sur les informations récoltées par ses voisins pour mener à bien sa tâche. C'est sur ce modèle de coopération que s'effectue toute la segmentation de l'image.



**Figure 1.14** Boucle de contrôle du comportement de perception. Ce comportement boucle sur une liste de pixels tant que celle-ci n'est pas vide.

## 1.9 Domaine d'application du traitement d'image :

Il existe plusieurs domaines d'application du traitement d'image, citons par exemple :

- La télédétection (météo, cartographie, analyse des ressources terrestres, astronomie).
- La médecine (radiographie, tomographie, les ultrasons, l'imagerie par résonance magnétique(IRM)).
- La Géologie, la chimie, la physique
- La photographie.

## Conclusion

Dans ce chapitre nous avons donné un aperçu rapide sur les définitions et les concepts de base concernant l'image, ces notions de base servent de fondement à la compréhension de différentes techniques de traitement d'images, Plusieurs méthodes classiques de traitement ont été proposés dans la littérature, nous avons présenté quelques unes qui nous semble les plus courantes dans le processus du traitement et analyse d'image.

Les prétraitements d'images permettent d'améliorer la qualité de l'image en vue de traitements ultérieurs. Deux grandes approches peuvent être envisagées pour extraire les zones pertinentes des images :

On cherche les discontinuités dans la scène, c'est l'approche contour.

On cherche des zones de niveaux de gris homogènes, c'est l'approche région.

A la fin de ce chapitre, nous avons cité les différentes Approches coopératives qui ont été élaborées pour le traitement d'images.

# Chapitre

*Evaluation de la qualité de  
l'image*

2



## Chapitre 02 : Évaluation de la qualité de l'image

### 2.1 Introduction :

Les photos, images ne bénéficient pas seulement d'un rôle d'accompagnement : ils attirent l'œil et facilitent la compréhension. L'image représente également le contenu le plus engageant, notamment sur les réseaux sociaux. En effet, les internautes ont tendance à enregistrer ou à partager les photos qu'ils aiment ou qui les touchent.

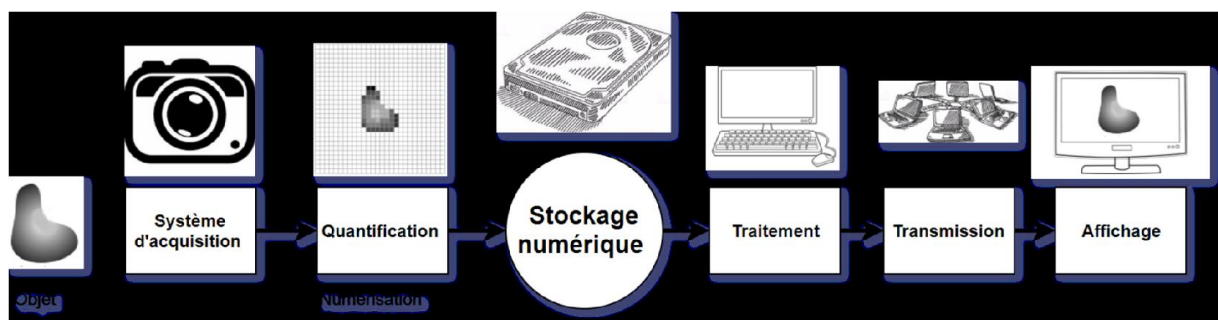
Aujourd'hui, créer un profil sur les réseaux sociaux (Facebook, Twitter,...), partager des souvenirs avec son entourage sont des utilisations banales de l'image. Ainsi, l'utilisation du contenu visuel a connu un fort développement qui tend à s'accroître dans les prochaines années. De ce fait, le domaine de l'image a connu un élargissement considérable à la fois en termes de matériel et de techniques de traitement d'images.

Depuis l'invention des capteurs CCD dispositif à transfert de charge (charge-coupled device) en 1969, l'évolution des moyens d'acquisition d'images ne cessent pas d'accroître. Sur le marché, il existe une multitude de choix d'appareils qui garantissent une capacité et une meilleure qualité d'image.

En parallèle, les techniques de traitement ont connu la même vitesse de progrès dans tout ce qui concerne la chaîne de traitement (acquisition, compression, transmission, diffusion...).

Au regard de cet essor majeur et la demande grandissante de l'utilisation des images, une meilleure qualité du contenu visuel est fortement exigée.

Durant la chaîne de traitement Figure 2.1 et à chaque niveau, l'image subit différentes dégradations/distorsions qui peuvent nuire à sa qualité perceptuelle.



**Figure 2.1:** Représentation des différentes phases de traitement d'une image

Étant donné que l'être humain est doté d'un système visuel sophistiqué capable d'évaluer la qualité perceptuelle de n'importe quelle image et afin de satisfaire les exigences de ce système la qualité du contenu visuel doit être contrôlée impérativement avant qu'elle arrive à ce dernier. Cela peut être réalisé avec différentes manières et il est appelé évaluation de la qualité perceptuelle.

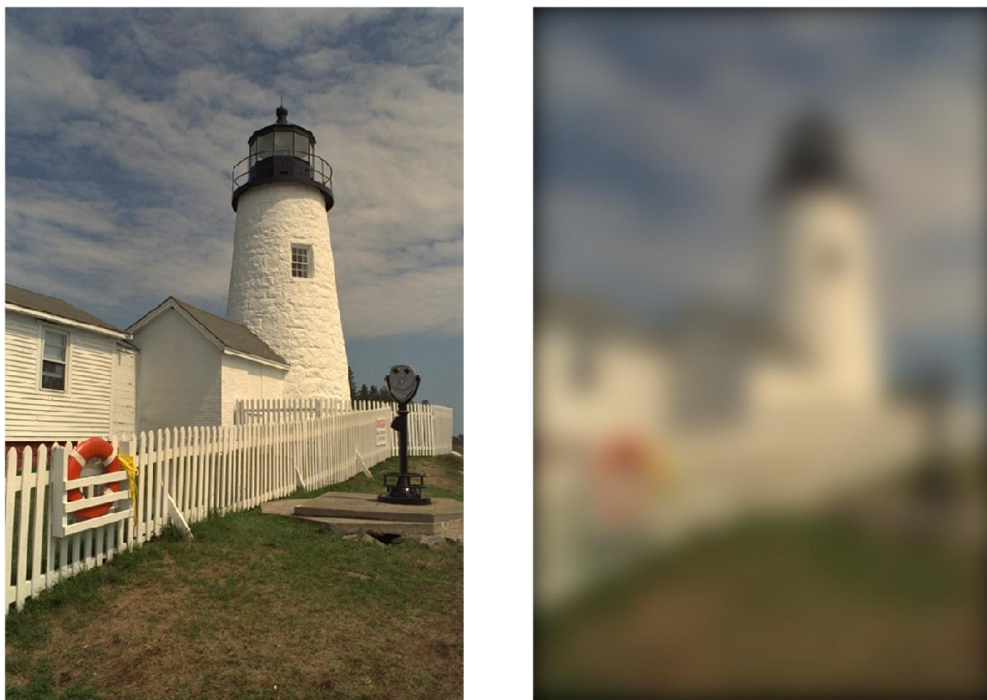
## 2.2 l'évaluation de la qualité d'image :

### 2.2.1 introduction :

La qualité d'image est un terme général employé par l'observateur pour évaluer la qualité inhérente de l'image. Elle est principalement subjective, signifiant que chaque observateur évalue la qualité d'image différemment. Certains peuvent penser qu'une image est d'assez haute qualité, mais on peut avoir un individu avec des niveaux plus élevés qui peut indiquer que l'image manque de qualité.

Évaluer la qualité d'une image c'est lui associer un ou plusieurs qualificatifs permettant de situer sa position relative dans un référentiel défini par notre sens et selon l'application envisagée. Le fait de capturer les défauts dans une image est appelé évaluation de qualité.

La Figure 2.2 représente deux images. L'une dite originale, il est supposé qu'elle ne contient pas de distorsion. La deuxième est une image déformée avec une dégradation. Notre système visuel humain SVH détecte facilement laquelle des deux images a la meilleure qualité.



**Figure 2.2:** L'effet de dégradation, (a)Originale, (b)Dégradée

Il existe deux types d'évaluation : **Subjective** et **objective**.

**-Évaluation subjective :** L'évaluation subjective fait appel à des observateurs humains (utilisateurs finaux) pour évaluer (ou comparer) la qualité d'une image (ou de plusieurs images) selon un protocole bien défini. Elle est considérée comme le moyen le plus fiable et donc la référence pour comparer les différentes métriques proposées.

- **l'évaluation objective** : Consiste à concevoir des modèles mathématiques capables de prédire la qualité d'une image avec précision et d'une façon automatique. La métrique objective devra pouvoir imiter les prévisions de qualité données par un observateur humain.

### 2.3 la fidélité:

La fidélité se réfère à la notion de conservation de l'information. Il s'agit de mesurer la similarité entre un stimulus original et une version à évaluer.

### 2.4 Différents types de dégradations :

On appelle distorsion ou dégradation tout défaut pertinent qui peut être vu, perçu ou remarqué dans une image.

Dans la pratique, il existe plusieurs types de dégradations. Les sources de distorsions se différencient, il y'a celles qui sont introduites durant la phase d'acquisition, d'autres sont dues aux traitements ... etc.

Il existe plusieurs types de dégradations qui sont :

#### 2.4.1 Dégradation de flou :

Effet esthétique qui donne à voir un contour imprécis. Ce type d'erreur qui apparaît principalement dans la compression JPEG et JPEG2000 d'où la perte de netteté de l'image. Cela se caractérise par une image plus floue, dont les bords des objets sont plus diffus.

Tout au long de la chaîne, on peut avoir différents types de flou.

- **Défocalisation (Out-of-focus) (Flou au niveau de la capture)** : Le flou peut provenir d'une erreur de manipulation du matériel d'acquisition telle qu'une mauvaise focalisation, le flou de défocalisation est lié à la notion de profondeur de champ qui n'est rien d'autre que l'espace d'excursion du plan de focalisation permettant d'avoir une image nette.

En effet, lors d'une prise de vue, la focalisation sur un plan donné entraîne inévitablement un flou au niveau des objets se trouvant en dehors du champ de netteté. Les images hors focus apparaissent floues.

La Figure 2.3 montre une défocalisation lors d'une prise de vue



Figure 2.3: Flou de défocalisation.

- **Flou de mouvement** : Le flou de mouvement dit aussi cinétique désigne le flou dû au mouvement rapide des objets photographiés pendant l'enregistrement. Le flou de mouvement est formé soit par le mouvement de l'objet quand l'appareil de capture est fixe, soit par la scène à filmer si l'appareil suit l'objet en mouvement.

La Figure 2.4 présente le flou de mouvement.



Figure 2.4 – Exemple de flou de mouvement.

- **Flou de perte d'information** : La perte d'information le long des canaux de transmission est aussi une cause d'apparition de flou dans l'image. La figure 2.5 présente un exemple d'image dont la perte de données est à l'origine du flou.



Figure 2.5 – Flou dû à une perte d'information.

- **Flou dû à la compression** : Les processus de compression représentent une source importante du flou. En général, des filtres passe-bas sont appliqués sur l'image, ils opèrent sur les hautes fréquences. Cela induit une perte de détails et de netteté. La Figure 2.6 représente les effets des deux compressions JPEG2000 et JPEG.



**Figure 2.6:** Flou dû à la compression, (a)Image compressée par le JPEG2000, (b)Image compressée par le JPEG.

- **Flou dû au traitement :** Certains traitements visant à atténuer quelques distorsions tels que le filtrage du bruit peuvent engendrer du flou. Un exemple d'apparition de flou résultant de l'application d'un filtrage passe bas est illustré par la figure 2.7



**Figure 2.7** – Un flou dû à un filtrage.

- **Flou dû à la transmission :** La transmission des images dans les canaux produit souvent des pertes d'informations, ces pertes se traduisent par des régions floues. La Figure 2.8 représente une image résultante de la simulation d'une transmission dans les canaux Rayleigh.



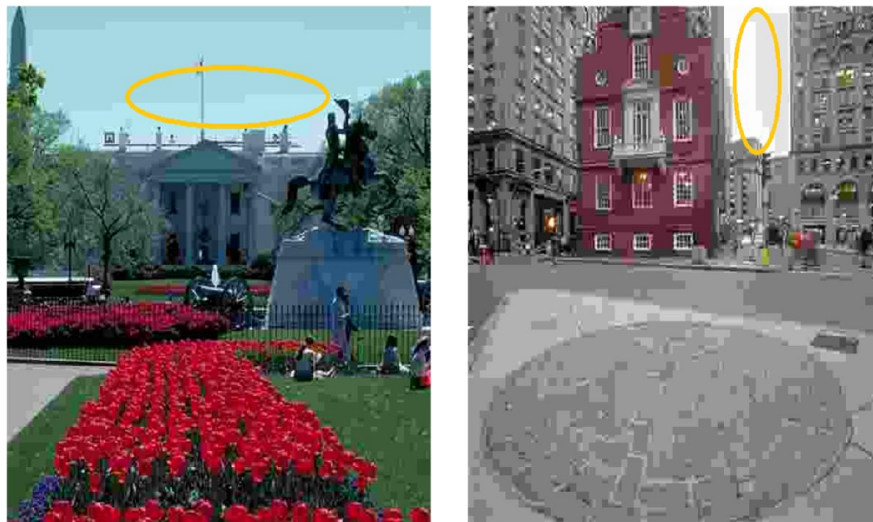
**Figure 2.8:** Flou dû à la transmission.

### 2.4.2 Effets de bloc :

C'est une dégradation où visuellement la distorsion se manifeste généralement au niveau des frontières entre blocs et apparaît comme des contours verticaux et horizontaux dont la visibilité dépend fortement de la distribution spatiale du signal image. La régularité de ces contours occasionne une gêne importante et rend alors cet artefact très visible.

De plus, la sensibilité accrue du SVH aux contrastes verticaux et horizontaux, et l'accroissement de la perception des contrastes entre deux régions adjacentes ne font qu'accroître la gêne occasionnée par cet artefact. La figure 2.9 illustre un exemple d'images dégradées.

Les effets de bloc sont la conséquence d'un découpage de l'image en blocs et leur traitement de façon indépendante. C'est généralement le cas des méthodes de compression par bloc telles que JPEG.



**Figure 2.9:** Images compressées par JPEG : (a) Effets de blocs horizontal,  
(b) Effets de blocs vertical.

### 2.4.3 Contraste :

Le contraste est un facteur déterminant dans la perception de la qualité visuelle. La distorsion du contraste est souvent introduite durant la phase d'acquisition. Les limitations du matériel d'acquisition et les conditions lumineuses sont les causes principales des pertes de contraste et de la visibilité des détails de l'image.

Il existe d'autres sources de la distorsion contraste. Parmi ces sources on peut citer le processus de rehaussement qui crée une perte ou une sur-contraste La Figure 2.10 montre l'effet de contraste sur la qualité perceptuelle.



Figure 2.10: Images représentant la dégradation contraste.

## 2.5 Évaluation subjective :

Considérée comme le moyen le plus fiable pour mesurer la qualité réelle, l'évaluation subjective fait appel directement à l'observateur humain. Ce dernier est amené à juger la qualité des images qui lui sont présentées selon une grille d'évaluation à plusieurs niveaux d'appréciation. À l'issue de ces tests, une note subjective appelée MOS (Mean Opinion Score) est obtenue.

Pendant la réalisation de ces tests, certains facteurs doivent être pris en charge affectant effectivement les jugements des observateurs, Ces facteurs sont énumérés dans ce qui suit.

- **Distance d'observation :** La distance d'observation affecte la visibilité d'un stimulus. Cette distance est fixée entre 4 à 6 fois la hauteur de l'image de projection et qu'il faut conserver tout au long du test.
- **Écran :** Il est nécessaire de calibrer le support d'affichage qu'est l'écran afin qu'il fonctionne dans des conditions optimales. Sans cette calibration, les couleurs affichées peuvent être différentes du stimulus d'origine.
- **Conditions de visualisation :** L'environnement et l'éclairage de la salle affecte effectivement la perception des stimuli (séquence de test). Un éclairage élevé risque de troubler l'observateur et ainsi pourrir sa perception. En revanche, la couleur de fond de la salle affecte l'apparence de certaines couleurs.
- **Le choix des images :** Un panel d'images de contenus visuels assez variés et le plus représentatif possible des scènes, liées à l'application visée, doit être utilisé.
- **Facteurs psychologiques :** Pour minimiser les effets liés à l'état psychologique de l'observateur, il convient de débiter les tests par une séance d'initiation permettant à l'observateur de mieux appréhender le test. Il est aussi important d'expliquer clairement l'objectif du test et le protocole à respecter sans pour autant influencer le jugement des participants. De plus, il est préférable de

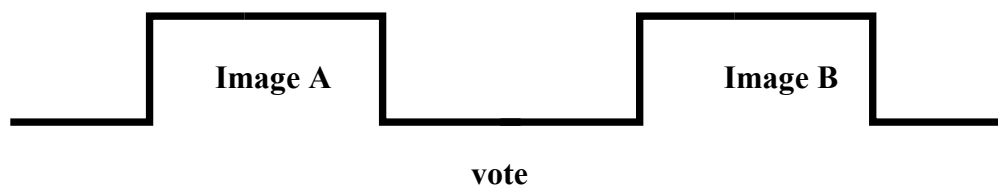
mener de nombreux tests de courte durée (environ 10min) et variés plutôt que de contraindre l'observateur à des séances longues (environ 30min) et fatigantes qui risquent de l'amener à un jugement rapide et aberrant.

- **Observateurs :** Il est recommandé d'avoir un panel d'observateurs le plus large possible, au moins quinze individus. Ils peuvent être experts ou novices, en ce qui concerne le thème de la campagne d'évaluation. Avant chaque séance, les observateurs seront sélectionnés pour leur acuité visuelle normale ou rendue normale par correction et leur vision normale des couleurs. L'observateur doit avoir ainsi une acuité visuelle de 10/10 pour les deux yeux avec ou sans correction. Le panel d'observateurs choisi dépend à la fois des types de traitement à évaluer et de la nature de l'étude à mener. Il faut donc prendre en compte, lors de la sélection de ce panel, plusieurs critères comme l'âge mais aussi l'origine socioculturelle, professionnelle.

### 2.5.1 Protocoles d'évaluation subjective :

Différentes méthodes d'évaluation subjective ont été définies par l'ITU, On recense essentiellement trois grandes familles : les tests à simple stimulus permettant de juger de la qualité d'image sans aucune information relative à l'image Originale(référence). Les tests à double stimuli où il est demandé à l'observateur de juger la similarité entre deux images et les tests comparatifs dont l'objectif principal est de comparer deux ou plusieurs stimuli. Ces méthodes sont décrites ci-après.

#### 2.5.1.1 Simple stimulus :



**Figure 2.11** – Méthode à simple stimulus.

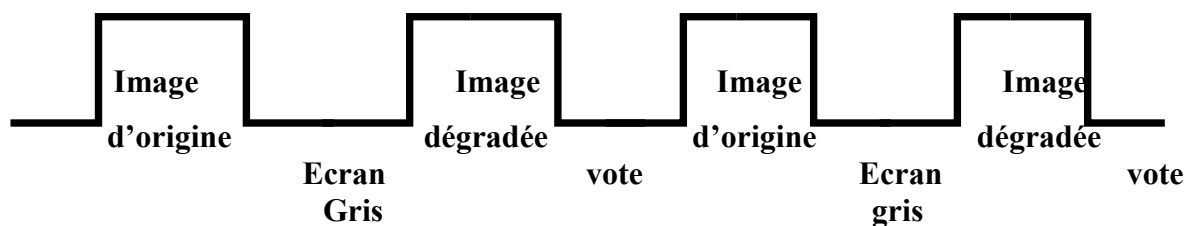
La méthode à simple stimulus permet de juger de la qualité d'un stimulus à la fois. Il est à noter que le mode de présentation, illustré par la figure 2.11, se voit ainsi limité. Les images sont présentées une par une avec un temps de latence entre deux présentations. Ce temps permet à l'observateur de noter la qualité de l'image.

Excellente
Bonne
Assez bonne
Médiocre
Mauvaise

**Table 2.1** – Echelle d'évaluation de qualité d'image.

La note Excellente indique une meilleure qualité. Ces échelles peuvent varier de 1 à 5 jusqu'à 100.

### 2.5.1.2 Double stimuli :



**Figure 2.12** – méthode à double stimulus.

Contrairement à la méthode à simple stimulus, cette méthode a pour objectif de mesurer la qualité (ou fidélité perceptuelle) d'un stimulus par rapport à sa version originale. L'image de référence est d'abord présentée, suivie d'un écran gris et de sa version dégradée (ou pas). Un second écran gris est affiché, il permet à l'observateur de disposer d'un temps de vote. Un exemple de présentation est illustré par la figure 2.12. Il est à noter que la durée de présentation doit être identique pour les deux stimuli.

Différentes échelles d'appréciation peuvent être utilisées où chaque note désigne la qualité ou le niveau de perception de la dégradation. Un exemple d'échelle de mesure à 5 niveaux est illustré par le tableau 2.2.

Notes	Qualité	Dégradation
5	Excellente	Imperceptible
4	Bonne	Perceptible
3	Passable	Peu dégradée
2	Médiocre	Dégradée
1	Mauvaise	Très dégradée

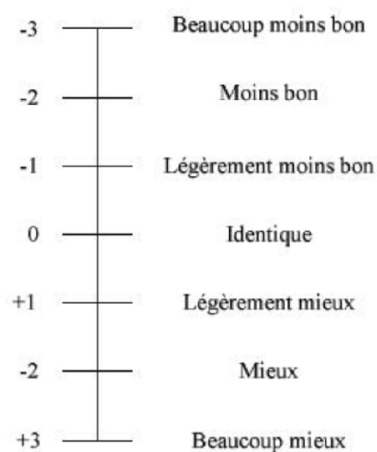
**Table 2.2** – Echelle d'évaluation de qualité d'image à 5 niveaux.

Beaucoup moins bon
Moins bon
Identique
Légèrement mieux
Mieux
Beaucoup mieux

**Table 2.3** – Echelle d'évaluation de la qualité

### 2.5.1.3 Méthode Comparatif :

Les méthodes comparatives d'évaluation de la qualité d'une image consistent à noter la différence entre deux versions d'une même image à l'aide d'une échelle : les deux images sont montrées en même temps et l'observateur juge laquelle est de meilleure qualité. Un exemple d'échelle comparative est donné à la figure 2.13.



**Figure. 2.13** – Une échelle comparative de qualité visuelle.

Une fois les tests réalisés, les résultats sont analysés et combinés de manière à obtenir une note unique par image décrivant sa qualité moyenne. Cette note appelée "Mean Opinion Score (MOS)" est donnée par la formule suivante :

$$\text{MOS}(i) = \frac{1}{\text{Nobs}} \sum_{j=1}^{\text{Nobs}} \text{Note}_i(j)$$

Où Nobs est le nombre total de participants et Note<sub>i</sub>(j) la note affectée à l'image *i* par l'observateur *j*.

### 2.5.2 Bases d'images dédiées à l'évaluation de performance :

Plusieurs bases d'images ont été proposées dans le domaine de la qualité d'images.

Dans cette partie nous présentons quelques bases d'images les plus utilisées dans la littérature. Pour chacune d'elles, nous présentons les différents types d'image ainsi que les protocoles utilisés dans l'évaluation subjective.

### 2.5.2.1 La base d'image LIVE :

Elle est construite par l'équipe du laboratoire d'image et de vidéo à l'université Austin Texas. Elle est constituée de 29 images de référence, qui ont permis de générer 779 images dégradées.

Les images de référence sont montrées dans la Figure 2.14. Elle utilise cinq différents types de dégradation:

dégradations dues aux compressions JPEG2000 (JP2K) et JPEG, bruit blanc Gaussien (WN), flou Gaussien (Gblur) et la dégradation due à la transmission dans les canaux Rayleigh (FF). Pour chaque image, un score subjectif DMOS est associé. Il représente la qualité de l'image dans un intervalle de 0 à 100 où 0 représente une mauvaise qualité et 100 représente une bonne qualité.

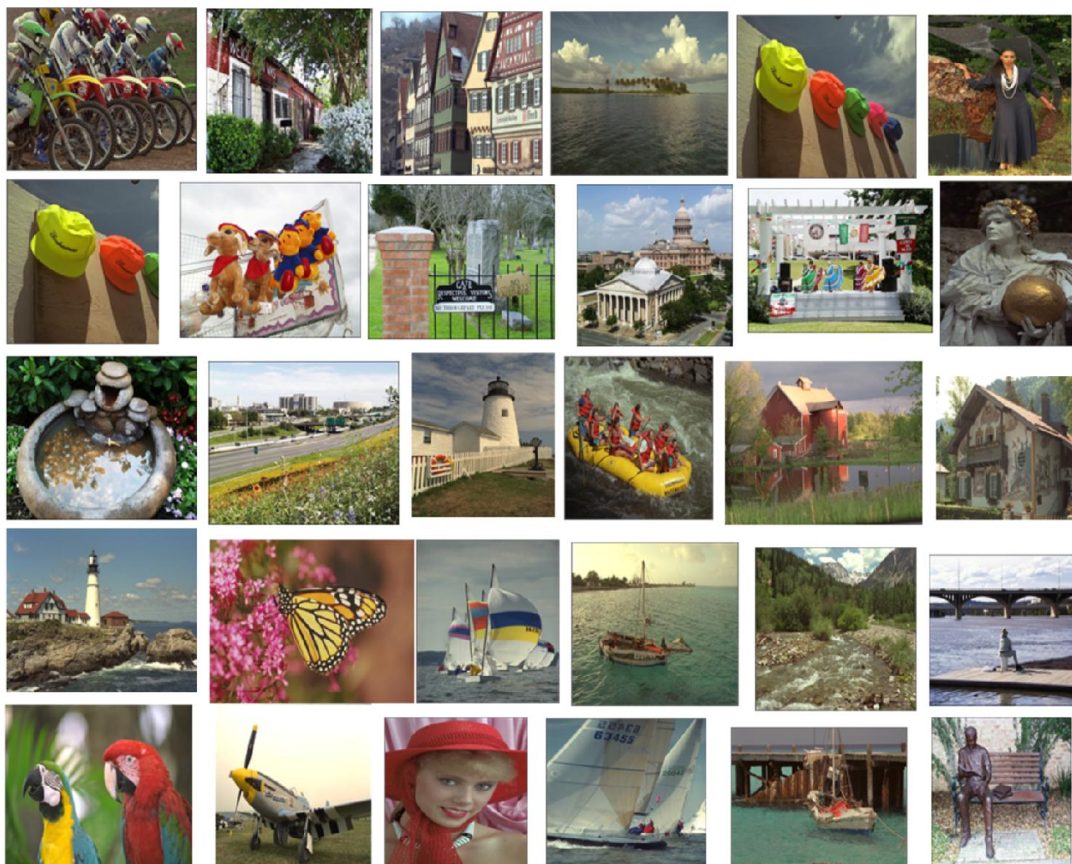


Figure 2.14: Images de référence de la base LIVE.

### 2.5.2.2 La base CSIQ :

La base CSIQ est composée de 30 images originales Figure 2.15 où chaque image est déformée en utilisant six types de distorsions à quatre ou cinq niveaux différents. Les distorsions utilisées dans CSIQ sont : la compression JPEG, la compression JPEG2000, la dégradation du contraste, bruit Gaussien additif et flou gaussien. Un total de 866 images déformées est généré. Les images CSIQ sont évaluées subjectivement et les notes sont présentées sous la forme de DMOS.

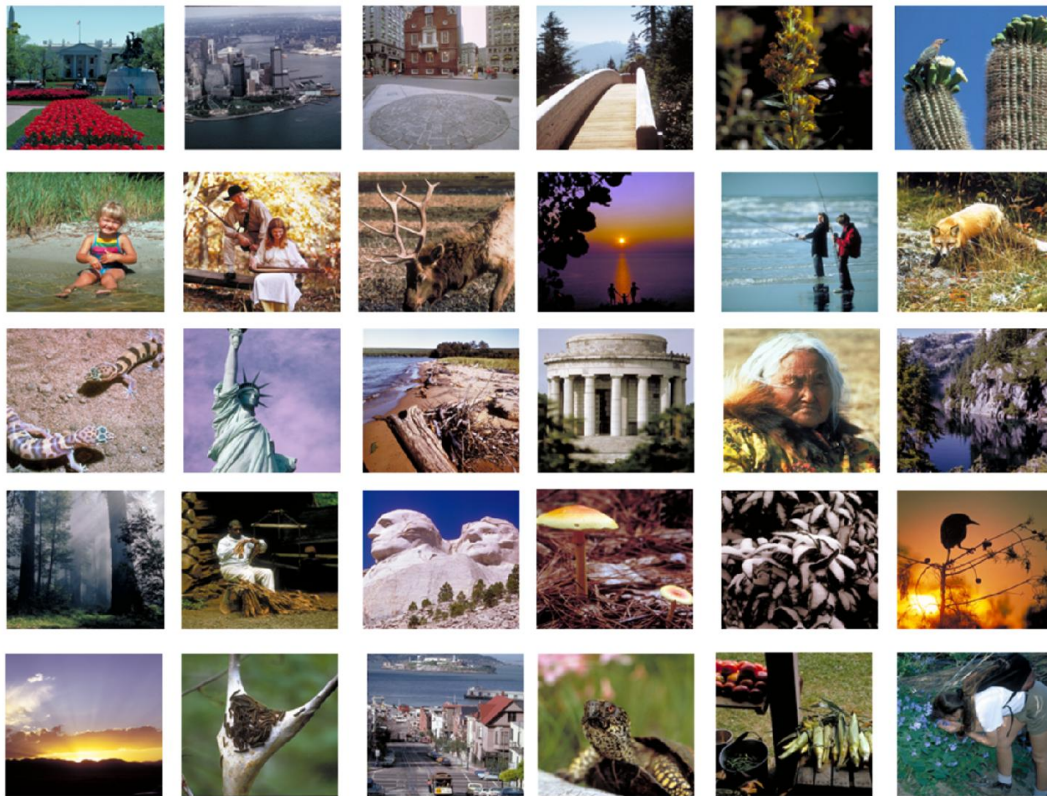


Figure 2.15: Images de référence de la base CSIQ.

### 2.5.2.3 La base d'image IVC :

Elle est générée à partir de 10 images originales voir la Figure 2.16. Les images dégradées sont obtenues en appliquant 4 algorithmes : JPEG, JPEG2000, LAR coding, Flou. Au final, 235 images ont été générées.

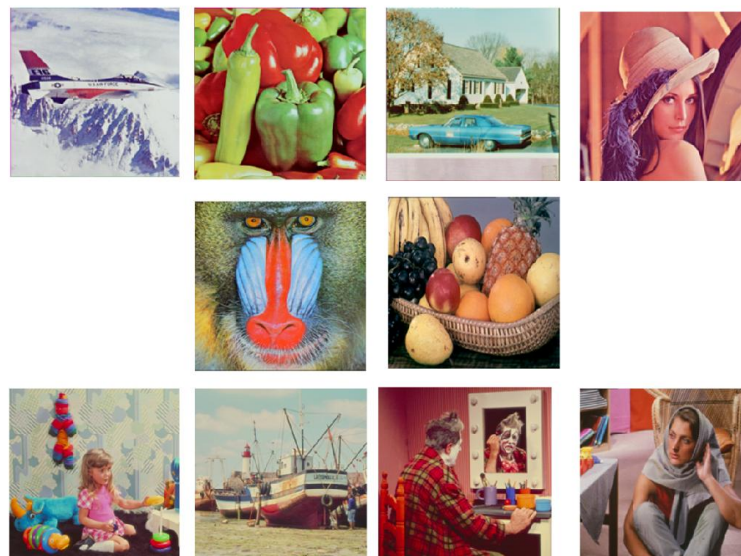


Figure 2.16: Images de référence de la base IVC.

### 2.5.2.4 La base Tampere TID :

Pour TID il existe deux bases, celle de 2008 et celle de 2013. La seule différence entre les deux bases c'est les niveaux des dégradations utilisées. Elles ont été constituées à partir de 25 images de référence voire la Figure 2.17. Avec 24 types de distorsions qui ont 4 (5) niveaux respectivement. Au total, il y a 1700 images dégradées pour la TID2008 et 3000 pour la TID2013. Les scores sont fournis en terme de MOS dans un intervalle de 0 à 9. La valeur la plus élevée de MOS (0 - minimal, 9 - maximal) correspond à une valeur de qualité visuelle supérieure.

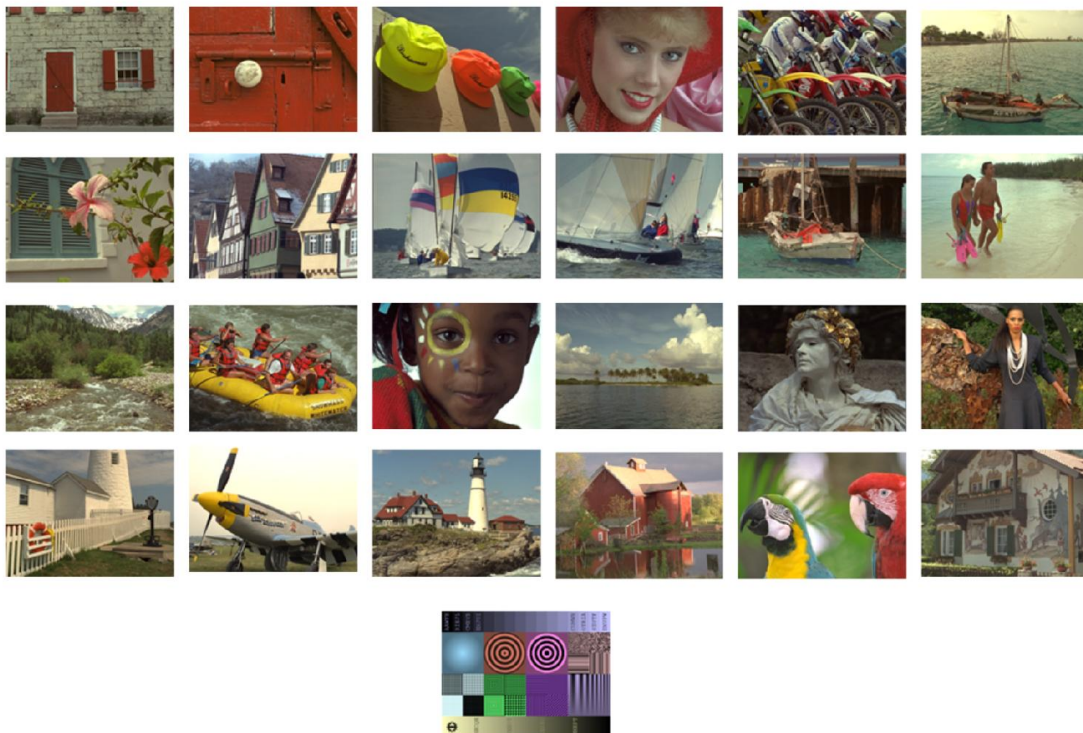
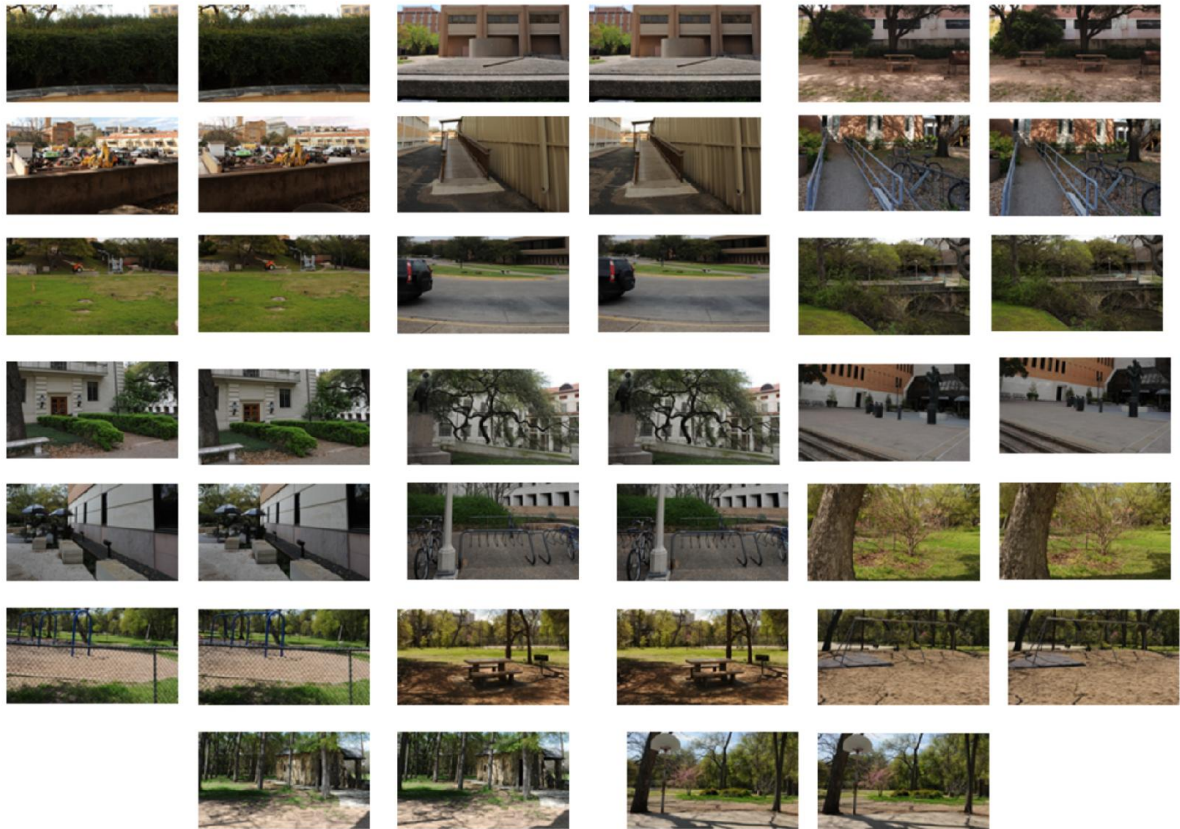


Figure 2.17: Images de référence de la base TID.

### 2.5.2.5 La base LIVE 3D phase I:

Elle est constituée de 20 images de références et 365 images dégradées. Il y a 80 images pour les dégradations JP2K, JPEG, WN, FF et 45 images pour la dégradation floue (Gblur). Pour chaque image, un score subjectif a été attribué sous la forme d'un DMOS. La Figure 2.18 présente les images de référence de cette base.



**Figure 2.18:** Images de référence de la base LIVE 3D phase I.

- Il existe d'autres bases d'images comme : La base LIVE 3D phase II, La base Waterloo-IVC 3D Phase I, La base Waterloo-IVC 3D phase II

## 2.6 Évaluation objective :

L'évaluation subjective reste la meilleure manière d'estimer la qualité visuelle. Le problème pour ce genre de méthodes est le coût en terme de temps et de ressource.

Cela a conduit les chercheurs à penser de proposer d'autres méthodes d'évaluation qui ont éventuellement les métriques objectives. Cette évaluation utilise des algorithmes. Pour évaluer la qualité d'une façon automatique.

Les méthodes d'évaluation de la qualité d'image peuvent être classifiées dans trois larges catégories :

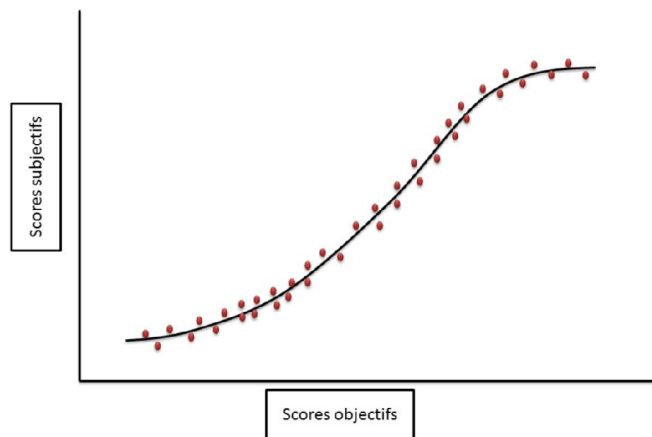
- **Les méthodes avec référence complète** : dans lesquelles l'algorithme a un accès à une version parfaite de l'image avec lesquelles il peut comparer la version dégradée. La version parfaite vient généralement d'un dispositif de haute qualité d'acquisition, après elle est dégradée par la compression et les erreurs de transmission.
- **Les méthodes sans référence**, dans lesquelles l'algorithme a un accès qu'au signal déformé et doit estimer la qualité du signal sans connaissance de la version parfaite. Puisque les méthodes

sans référence n'exigent aucune information de référence, elles peuvent être employées dans n'importe quelle application où la mesure de qualité est exigée.

- **Méthodes avec référence réduites**, dans lesquelles l'information partielle concernant la version parfaite est disponible. Un côté canal existe par lequel quelque information concernant la référence peut être rendue disponible à l'algorithme d'évaluation de la qualité. Les algorithmes réduits de référence emploient cette information de référence partielle pour juger la qualité du signal déformé.

Pour connaître l'efficacité de toute métrique, il doit être évalué, les scores obtenus par l'évaluation objective doivent être corrélés avec les scores subjectifs de manière prévisible et répétitive.

La figure 2.19 montre relation entre les scores objectifs et subjectifs.



**Figure 2.19:** Exemple de relation entre les scores objectifs et subjectifs.

# Chapitre

*les images et vidéos  
stéréoscopiques*

3



## Chapitre 03 : les images et vidéos stéréoscopiques

### 3.1-Introduction :

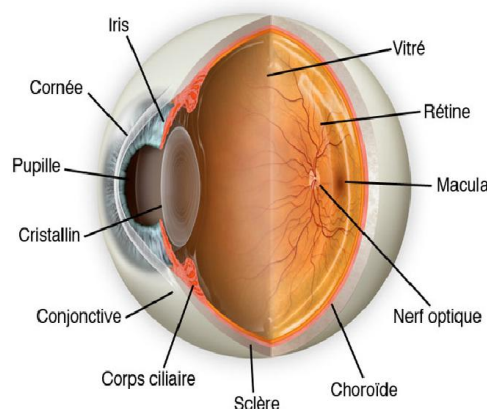
Le terme 2D signifie bidimensionnel, tandis que 3D signifie tridimensionnel. 2D représente un objet en seulement deux dimensions, tandis que 3D le représente en trois dimensions. Il vaut mieux y penser avec un exemple. Une voiture a trois dimensions. En fait, tout dans notre réalité est tridimensionnel. Il a une longueur, une largeur et une hauteur; cependant une photographie d'une voiture ou un dessin d'une voiture n'a que deux dimensions. Sur la photo ou le dessin, on ne peut voir et mesurer que la longueur et la hauteur, alors que la profondeur ou la largeur est indiscernable.

De nos jours, le contenu 3D devient très utile dans différents domaines : vidéos mobiles 3D, production des films 3D, outils de capture 3D et de visualisation, les performances de 3D augmentent de jour après jour, ce qui permet de visualiser des scènes 3D complexes en temps réel.

Dans ce chapitre on va présenter une petite image historique pour les stéréoscopes, des définitions de la vision humaine et la relation entre le cerveau et les yeux et en fin on va définir la stéréoscopie.

### 3.2-La vision humaine tridimensionnelle:

L'œil humain est simplement constitué. Il est composé d'un trou laissant passer la lumière appelé la pupille. Ensuite se trouve une lentille, qui fait converger les rayons lumineux pour que ceux-ci arrivent sur la rétine, sorte d'écrans qui capture la lumière et la transforme en impulsions nerveuses. Ces impulsions sont transmises au cerveau grâce aux nerf optiques, et les informations des deux yeux sont traitées dans les 5 zones visuelles (numérotée de V1 à V5). Dans ces zones, les formes, les couleurs et les mouvements sont analysés.

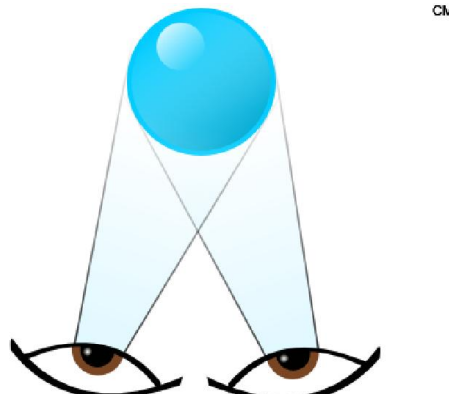


**Figure 3.1 : L'œil humain**

Le cerveau humain a la capacité de reconstituer une image en relief grâce aux informations en provenance de l'œil. En effet, l'humain possède deux yeux, donc deux informations différentes sont

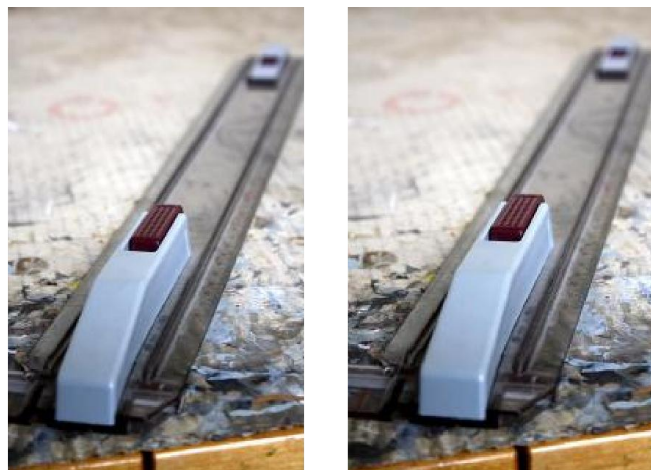
reçues pas le cerveau. Ces deux images mises en relation permettent d'apprécier la distance et le volume des objets observés. C'est ce qu'on appelle la vision tridimensionnelle. Cette appréciation est fiable seulement lorsque les objets se situaient à moins de 10 à 15 mètres. Au-delà de cette distance, la vision n'est plus aussi précise, et le cerveau utilise son expérience pour reconstituer le volume du paysage. Il se sert des objets de tailles connues, et de la superposition des différents composants de l'image pour apprécier le volume et la distance.

Un seul œil ne voit qu'en deux dimensions, maintenant pourquoi deux yeux nous permettent de voir en trois dimensions ? C'est en réalité parce que nous percevons deux images, une capturée par chaque œil. Et bien sûr ces deux images sont différentes. Si vous fermez un œil puis l'autre, vous verrez que certains objets bougent ou pivotent (en fait, ils pivotent et bougent tous, mais de manière plus ou moins perceptible). figure 3.2



**Figure 3.2 :** capture de l'image par les deux yeux

Et voici, deux images capturées par les deux yeux, avant traitement par le cerveau.

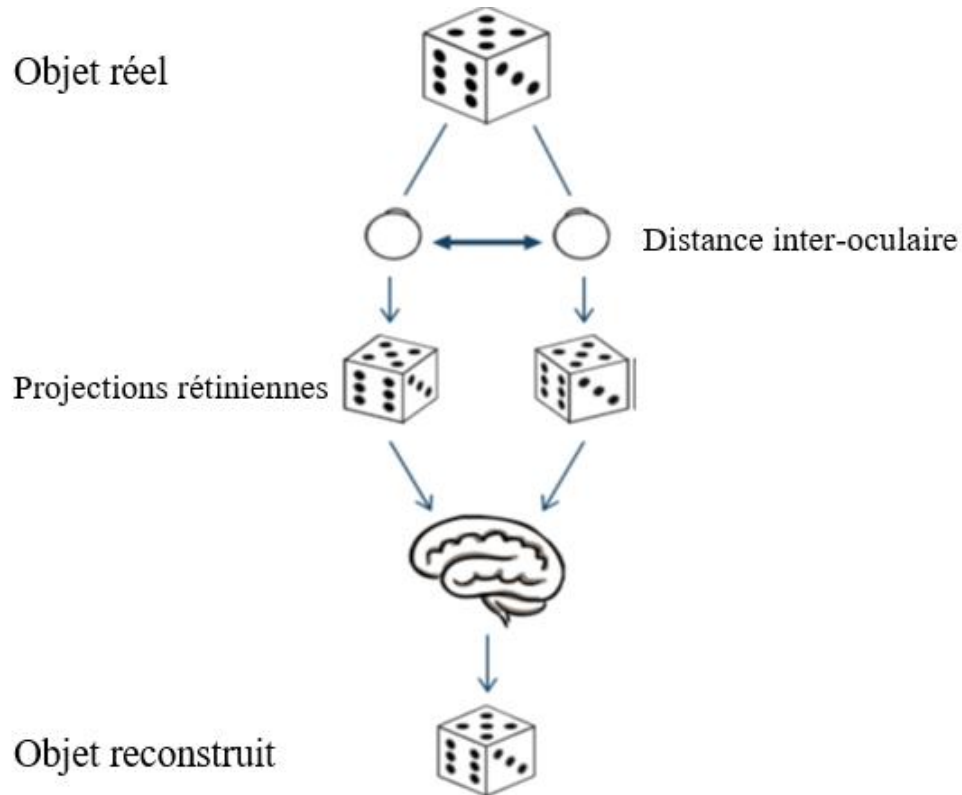


**Figure 3.3 :** deux images capturées par les deux yeux

Maintenant que nos images sont capturées, elles sont directement envoyées à traiter à notre cerveau. Que va faire celui-ci ? Il va tout simplement, à partir de la différence des deux images (différence de rotation, de déplacement, etc. des objets), calculer une troisième dimension. Donc, pour créer une image en 3D, il nous faut deux images différentes et un cerveau de tel sorte que chaque œil

envoie une image différente au cerveau pour que celui-ci crée la 3ème dimension. (Image tridimensionnelle).

Et voilà la figure 3.4 qui monte le fonctionnement de la vision humaine.



**Figure 3.4 :** une vision 3D de l'humain

### 3.3 - la Stéréoscopie :

#### 3.3.1- la vue en relief :

La vue en relief est notre vue naturelle, lorsque l'on regarde un objet on le voit en trois dimensions. Cette vue peut s'effectuer grâce à l'action de notre cerveau; il superpose les images vues par notre œil gauche et celles vues par notre œil droit, ce qui va donner lieu à notre vue en relief.

La sensation de relief dans la vision provient de ce que les deux yeux ne voient pas le même objet sous le même angle.

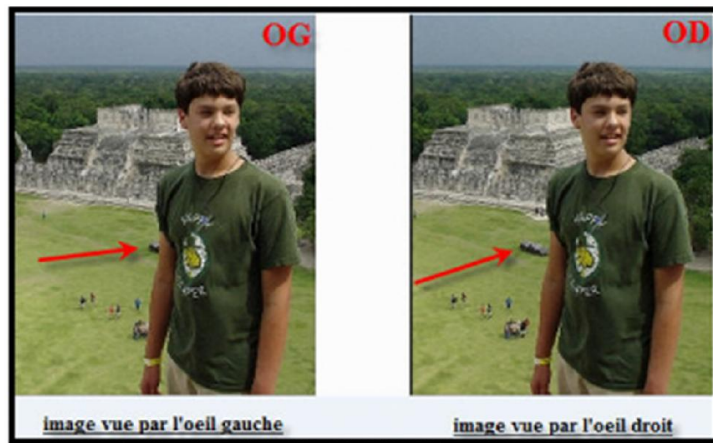
#### 3.4 c'est quoi la stéréoscopie :

La stéréoscopie est l'ensemble des techniques utilisées pour avoir, à partir de deux images en deux dimensions, une troisième image en 3 dimension.

La stéréoscopie comprend donc toutes les méthodes qui permettent d'obtenir une impression de relief, que ce soit en observant un objet à travers un instrument d'optique ou que ce soit en restituant une seule image en relief à partir de deux photographies.

Petit rappel d'anatomie : nous avons deux yeux séparés d'environ 65mm et qui voient chacun une image légèrement différente. Vous pouvez le vérifier en mettant votre doigt entre vos yeux et le paysage.

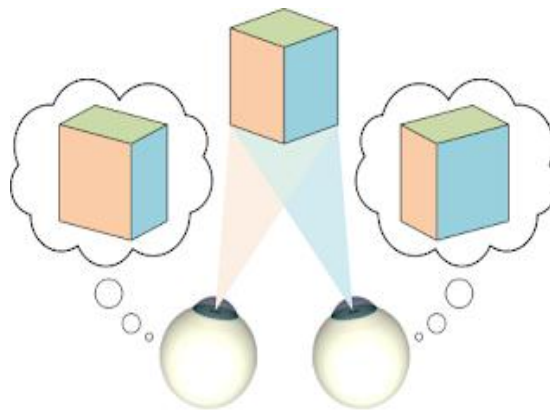
Regardez ce paysage en fermant un œil après l'autre. Vous aurez l'impression que votre doigt bouge



**Figure 3.5 :** Ces deux images sont transférées au cerveau, qui en « fabrique » une troisième en relief (3D).

On voit bien, par exemple, que la position de la voiture par rapport à la personne n'est pas la même pour les deux yeux. Sur l'image vue par l'œil gauche, on voit les objets de premier plan décalés vers la droite et sur l'image vue par l'œil droit ils sont décalés vers la gauche.

Nos deux yeux ne voient donc jamais la même chose ! A partir de deux images légèrement différentes, le cerveau parvient à reconstituer une scène en relief ! Le relief est donc produit à partir des différences entre les images vues par chacun des deux yeux.



**Figure 3.6 :** Principe de la vision stéréoscopie

Tout l'enjeu de la stéréoscopie est de produire les 2 images afin de "simuler les capteurs" humains.

### 3.5 l'Histoire de la Stéréoscopie :

Au 3ème siècle avant J.C, le géomètre grec Euclide, définit ainsi le principe de la vision en 3 dimensions : « Voir le relief, c'est recevoir au moyen de chaque œil l'impression simultanée de deux images dissemblables du même sujet ».

La vraie histoire de la stéréoscopie commence en 1832 avec l'Anglais Charles Wheatstone. Il étudie des moyens pour observer des dessins en relief. En 1838 il fait breveter le premier « stéréoscope ». C'est un appareil muni de deux miroirs dans lesquels on observe les dessins placés aux extrémités du stéréoscope.

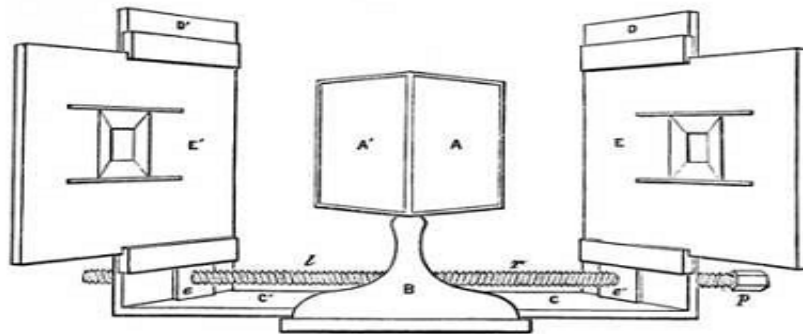


Figure 3.7 : Stéréoscope de Wheatstone

Quelques temps plus tard, Daguerre (à Paris) et Fox Talbot (à Londres) mettent au point des procédés de photographie. En 1840 Wheatstone demande à Talbot de prendre des photos en relief qu'il visionne en suite avec son stéréoscope.

En 1849, un autre Anglais, David Brewster invente un nouveau stéréoscope à lentilles. Jules Duboscq de Paris s'occupe de le construire. Il montre son invention lors de l'Exposition Universelle à Londres en 1851. La Reine Victoria s'y intéresse et la stéréoscopie devient à la mode.



Figure 3.8 : Stéréoscope de Jules Duboscq

Les photographes réalisent des couples stéréo de toutes sortes de sujets : nature, personnes, animaux, cérémonies, etc. En 1862 le principal distributeur de plaques stéréoscopiques en vend plus d'un million.

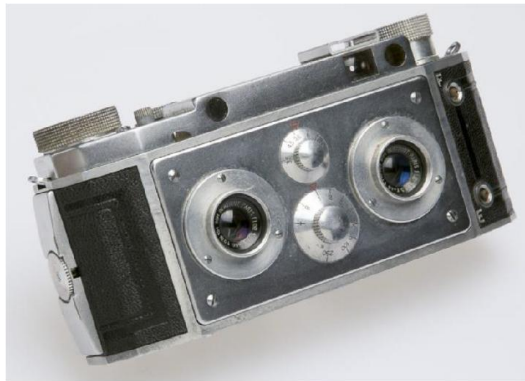
En 1893, Jules Richard fabrique le Verascope : premier appareil photo stéréoscopique rechargeable en plein jour.



**Figure 3.9** : Verascope

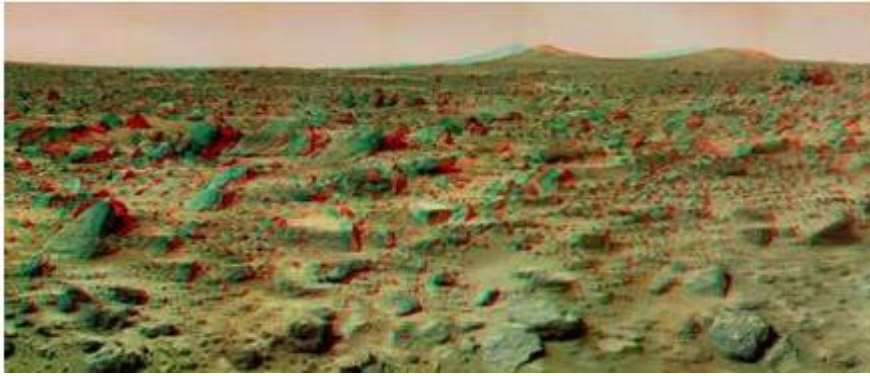
#### **Utilisation du support film pour les photos stéréoscopiques :**

Le Vériascope 40 de Jules Richard, commercialisé en 1939 est le premier appareil à film 24x36 grand public.



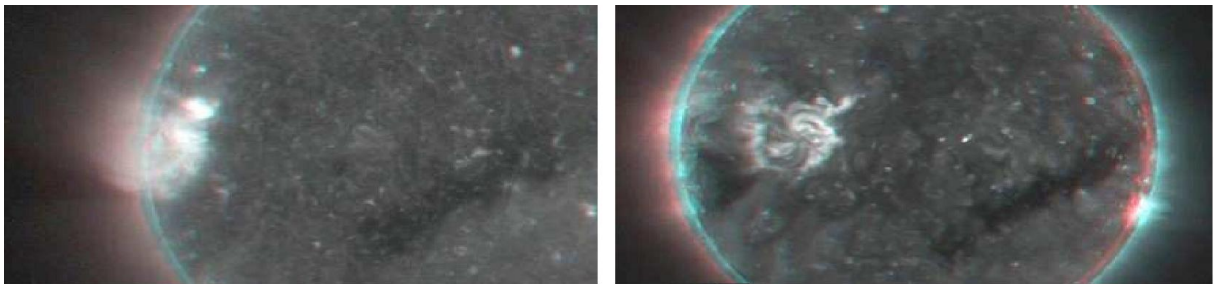
**Figure 3.10** Appareil à film 24x36, Vériascope 40 de Jules Richard

La NASA utilise beaucoup la stéréoscopie : des photos ont été prises par les astronautes lors des missions Apollo sur la Lune, des satellites ont photographié des planètes du système solaire en relief, la mission Mars Pathfinder a amené des images du sol martien en 3 dimensions :



**Figure 3.11** ; Photo prise sur mars

Depuis 2006, la NASA fait des photographies en relief du soleil. Elle a envoyé deux satellites qui se séparent de la terre dans des directions différentes afin d’avoir un écartement de plus en plus grand en visant le soleil.



**Figure 3.12:** Photo du soleil

De nos jours, la 3D est revenue en force grâce au cinéma et principalement par le film «Avatar» de James Cameron



**Figure 3.13** : Avatar



Figure 3.14 : Caméra 3D Professionnelle

### 3.6 Les différents types de stéréoscope :

Les stéréoscopes se sont des visionneuses qui permettent à chaque œil de ne voir que l'image qui lui est destinée (image gauche à l'œil gauche, image droite à l'œil droit) afin de reconstituer l'image en relief. Il en existe plusieurs types différents : à miroirs, à lentilles ou à prismes.

#### 3.6.1 stéréoscope de miroirs :

Inventé en 1838 par Charles Wheatstone, est muni de deux miroirs à 45° qui reflètent les images placées à chaque extrémité du stéréoscope. En regardant dans ces miroirs, une nouvelle image se forme en relief face à l'utilisateur

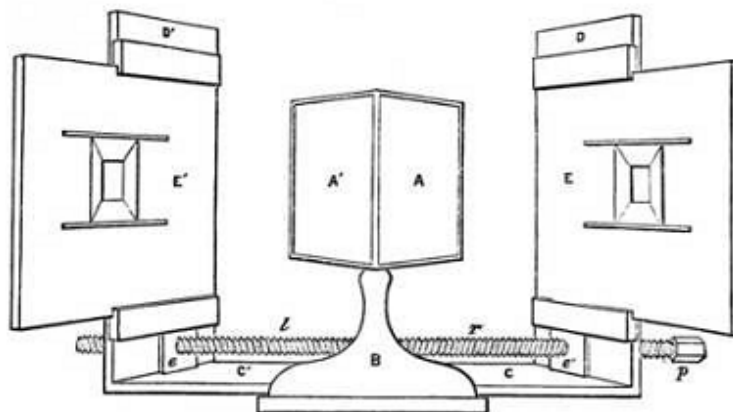
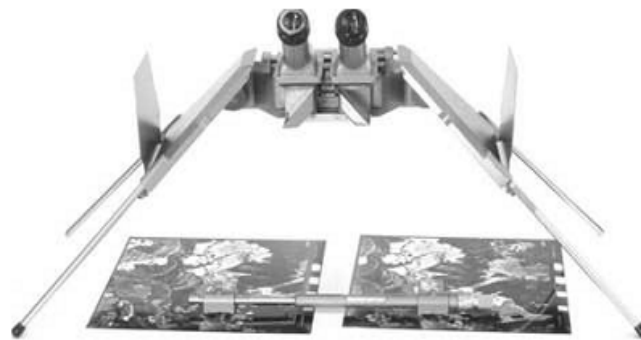


Figure 3.15: Stéréoscope de Wheatstone

Ce système est toujours utilisé de nos jours, principalement pour la cartographie, mais avec deux miroirs en plus. Ceux-ci permettent de voir de grandes images l'une à côté de l'autre.



**Figure 3.16 :** Cartographie

### 3.6.2 stéréoscope à lentilles:

En 1849, Sir David Brewster fait breveter un stéréoscope à lentilles. Les photos sont installées côte à côte, mais leur largeur ne doit pas dépasser l'écart entre les yeux. Le format 6 x 13 cm (dimension totale des 2 images) est celui qui convient le mieux.



**Figure 3.17 :** stéréoscope a lentilles

Ce type de visionneuse est souvent fabriqué soit avec un fond en verre dépoli mobile, soit avec une partie du dessus qui s'ouvre. Ceci permet de regarder des photos en verre ou imprimées sur carton. Nous utilisons aussi ce type de stéréoscope pour visionner les diapositives ou les photos papier. Ils sont généralement assez petits et même parfois pliables.

### 3.6.3 stéréoscope à prismes :

En 1859, Oliver Wendell Holmes met au point un stéréoscope plus léger et meilleur marché que tous les autres modèles de l'époque. Il est muni de deux prismes qui permettent de regarder des

images d'un format 9 x 18 cm au total. L'utilisateur ne peut voir avec chaque œil que l'image appropriée car une paroi verticale de séparation interdit à son regard de s'égarer. De plus, son front prend appui sur une sorte d'ocillère qui lui permet d'être stable et de ne pas avoir de lumières parasites dans les yeux.



Figure 3.18 : stéréoscope de Holmes

### 3.6.4 l'ère du numérique :

De nos jours, c'est-à-dire à l'ère du numérique, il existe des visionneuses qui permettent de regarder les images 3D sur les écrans d'ordinateurs (le Pokescope par exemple).



Figure 3.19 : Pokescope

Différentes marques d'écrans proposent des modèles capables de travailler directement en relief. Ils sont utilisables avec des lunettes actives de chez Nvidia. Certaines marques proposent aussi des écrans utilisables avec des lunettes passives.



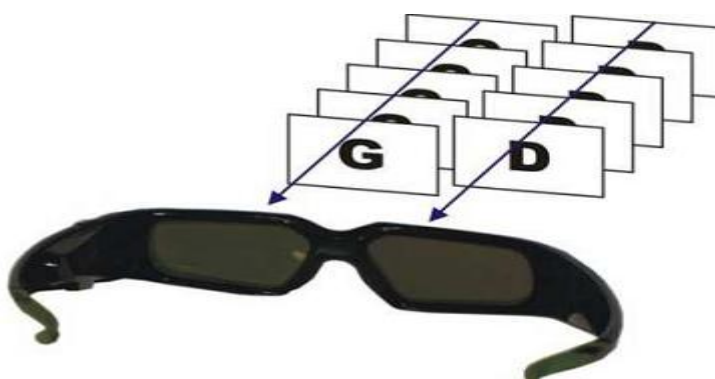
**Figure 3.20:** Lunettes actives Nvidia avec écran Samsung

Pour une plus grande quantité de spectateurs, il est possible de faire de la projection en utilisant des projecteurs de données et des lunettes passives.

En 2010, une norme de haute-définition 3D (HD-3D) a été décidée et certains fabricants font des téléviseurs capables de gérer la 3D. Des disques Blu-ray sortent en 3D et sont utilisables sur ces téléviseurs.

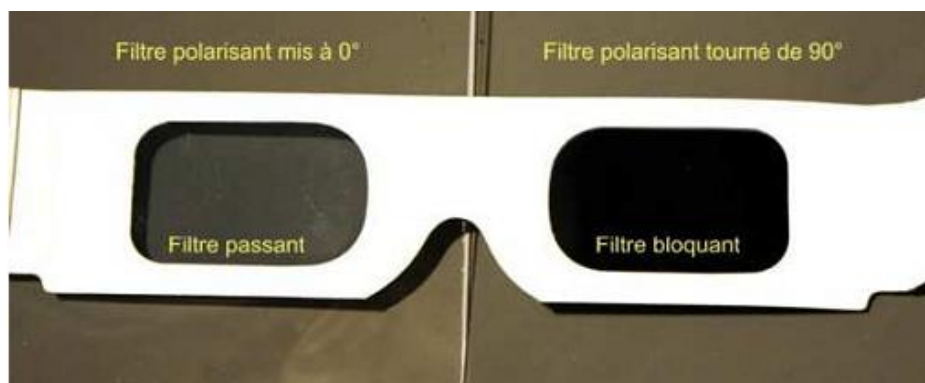
Les lunettes utilisées sont soit actives, soit passives :

Lunettes actives: elles sont équipées de petits écrans LCD qui sont soit transparents soit complètement opaques, en synchronisation avec les images gauche et droite diffusées alternativement. Elles ont une petite électronique embarquée avec une pile ou un accumulateur qu'il faut recharger.



**Figure 3.21** Lunettes actives (ou à LCD)

Lunettes passives : ce sont des lunettes simples avec des filtres polarisants. Elles sont aussi utilisées pour la projection. Des filtres polarisants décalés de  $90^\circ$  l'un par rapport à l'autre, sont placés devant les projecteurs. Ils cachent l'image gauche à l'œil droit et l'image droite à l'œil gauche.



**Figure 3.22** : Lunettes passives (ou polarisantes)

Pour les télévisions, la tendance actuelle va vers les lunettes passives. Elles sont généralement plus légères que les actives et demandent moins d'entretien.

### 3.7 Comment faire de la Stéréoscopie ?

Pour faire de la photographie en relief, il suffit de prendre deux photos légèrement décalées. Chaque œil regarde ensuite l'image qui lui correspond et le cerveau reconstitue le relief.

La méthode la plus simple est de prendre un appareil photo numérique standard, de bien se caler sur les pieds, de faire une première photo en prenant appui sur une jambe, puis, sans bouger les pieds, de prendre appui sur l'autre jambe et de faire la deuxième photo. Cette méthode impose de ne prendre en photo que des objets statiques. Ceux en mouvements ne seront pas au même endroit sur les deux images et donc impossible à regarder en relief.



*Prendre appui sur une jambe  
et faire une première photo*

Si vous avez la possibilité d'avoir deux mêmes appareils photo, fixez-les l'un à côté de l'autre sur une petite planche, bricolez un système qui permette de déclencher les deux appareils en même temps et prenez les photos. Il existe des systèmes de synchronisation tels que le logiciel *StereoData Maker* qui fonctionnent avec certains appareils.



**Figure 3.23** Canon montés ensemble et synchronisés par le logiciel StereoData Maker

Par ces méthodes, c'est-à-dire en ayant un écartement entre les deux prises de vues à peu près équivalent à celui des yeux (environ 65mm), le relief sera le même que celui que nous voyons dans la réalité.

### 3.8 Les images stéréoscopiques :

Les images stéréoscopiques sont constituées de deux vues de la même scène, présentées côte à côte. L'une correspond à l'œil droit, l'autre à l'œil gauche. Différents procédés permettent au cerveau de l'observateur de fusionner les deux images reçues séparément par l'œil droit et par l'œil gauche et de percevoir ainsi le relief et ça donner un effet de relief.

Cet effet de relief est produit par la parallaxe. La parallaxe est l'écart entre les deux positions, sur le support de l'image, des vues gauche et droite d'un même objet. Les deux caméras doivent être ajustées pour que la parallaxe correspondant à un point à l'arrière-plan ne dépasse pas 65 mm, c'est-à-dire la distance moyenne entre les pupilles d'un adulte.

### 3.9 Comment créer une image 3D stéréoscopique ?

la figure 3.24 suivante montre la façon qui nous permet de créer les images stéréoscopiques 3D

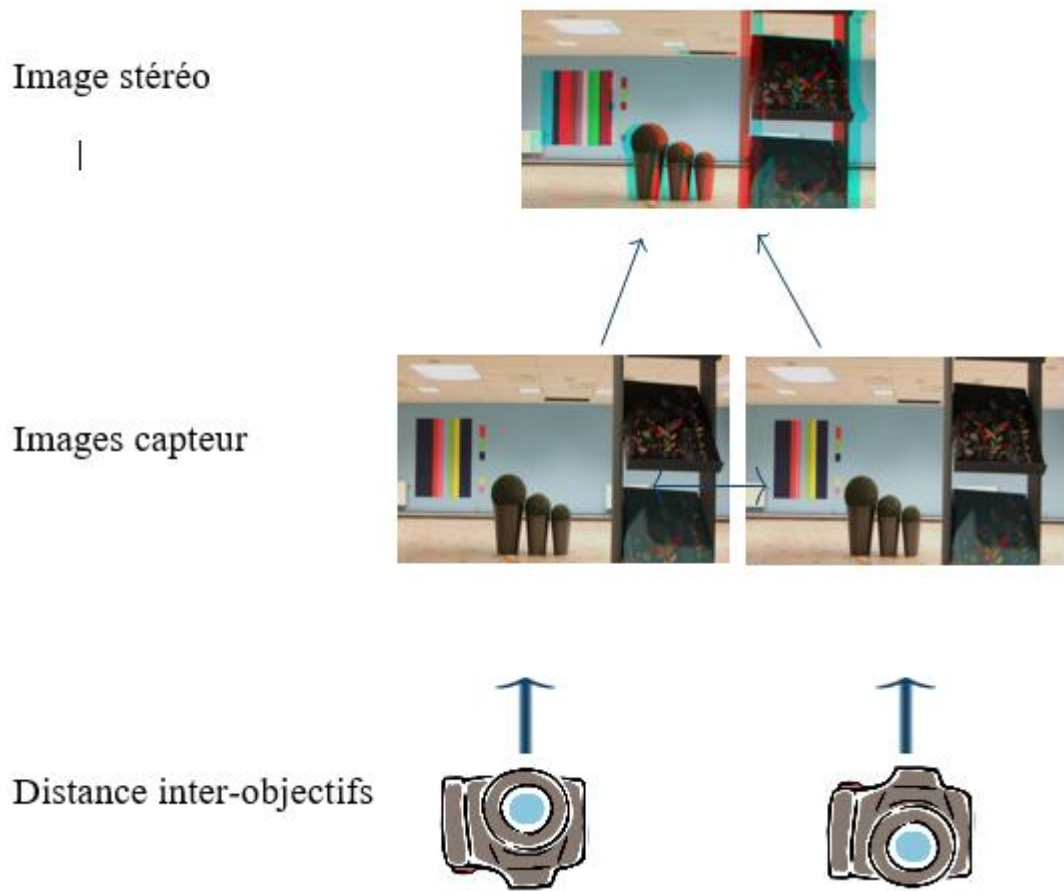


Figure 3.24 constitution d'images stéréoscopiques 3D

# Chapitre

*Évaluation de la qualité  
d'image et vidéo stéréoscopique  
sans référence basée sur les  
entropies spatiales et spectrales  
de l'image*

4



## **Chapitre 4 : Évaluation de la qualité d'image et vidéo stéréoscopique sans référence basée sur les entropies spatiales et spectrales de l'image**

### **Introduction :**

L'omniprésence de l'information visuelle (sous forme d'images et de vidéos) dans tous les secteurs a besoin de méthodes fiables et efficaces pour évaluer sa qualité. Par conséquent, de nombreux chercheurs se sont concentrés à développer des métriques de qualité. L'évaluation de la qualité de l'image (EQI) peut être organisée en trois grandes classes. La première classe comporte les métriques avec référence complète (MAR-EQI). Pour cette classe, l'image de référence est utilisée avec l'image dégradée, où on cherche à quantifier la moindre différence entre les deux. Cette quantification est transformée à un indice de qualité.

La deuxième classe est l'évaluation avec référence réduite (MRR-EQI). Pour celle-ci nous n'avons pas accès à la globalité de l'image de référence, mais plutôt à une partie. Selon ce qu'on possède de la référence, les mêmes informations sont extraites de l'image dégradée. Une comparaison est effectuée pour évaluer la qualité de l'image en question.

La troisième classe est dédiée à l'évaluation sans référence (MSR-EQI). Dans ce type d'évaluation, on ne possède que l'image dégradée. Cela rend plus difficile ce genre d'évaluation mais elle reste la plus recherchée. Ceci est due au fait que dans la majorité des applications l'image de référence est inexistante.

Dans ce chapitre, nous allons proposer une métrique sans référence basé sur les entropies spatiales et spectrales de l'image pour l'évaluation de qualité des images (des vidéos) stéréoscopique 3D affectées par des différentes distorsions, où il a été incorporé la vision binoculaire dans le processus d'évaluation, les deux vue gauche et droite de l'image (des vidéos) S3D sont combiné par un modèle de combinaison binoculaire, (l'activité spatiale et binoculaire) et l'image de produit sont calculées. Cette méthode utilise des caractéristique d'entropie spatiale et spectrale à partir d'une image (des vidéos) dégradée.

Et pour dériver ou prédire le score de la qualité perceptuelle, une régression à l'aide de machines à vecteurs de supports (SVM) (SVR) est utilisée.

80% de la base d'images (des vidéos) S3D est utilisé par le SVR pour la phase d'apprentissage qui est cruciale pour les machines à vecteur de support. Cette étape permet d'entraîner la machine à reconnaître comment chaque distorsion affecte l'image, ce qui produit un modèle d'évaluation de qualité d'image.

20% de la base d'images (des vidéos) est exploité par le modèle réalisé (implémenté) pour le test.

Ces résultats de tests (scores objectifs) sont comparés avec le DIMOS (scores subjectifs) pour évaluer la performance de cette méthode.

## 4.1 machines à vecteurs de support (SVM)

### 4.1.1 Introduction :

Les travaux de V. Vapnik, publiés en 1979, sur la théorie statistique de l'apprentissage peuvent être considérés comme le point de départ de ce qui devint par la suite les machines à vecteurs supports (ou SVM pour Support Vector Machines). L'intérêt pour ce sujet ne prit de l'ampleur qu'au début des années 1990, en réponse à différents problèmes d'apprentissage supervisé. Vapnik et Chervonenkis avaient introduit en 1971, dans le domaine de la classification, une manière de quantifier la capacité d'apprentissage d'une famille de fonctions en définissant une quantité (appelée depuis dimension de Vapnik-Chervonenkis ou VC-dimension) définie comme le nombre maximum d'exemples séparables par un classifieur à deux classes issu de cette famille.

Les machines à vecteurs supports ont été inventées en 1992, mais le terme « machines à vecteurs supports » n'est apparu qu'en 1995.

Dans cette partie, nous introduirons les notions sur l'apprentissage statistiques et le principe de machine à vecteur support et l'apprentissage dans le cas de classification. et enfin on aborde un cas particulier de la méthode SVM, SVM pour la régression SVR.

### 4.1.2 Notions sur l'apprentissage :

D'après Mari & Napoli (1996) une classification, c'est mettre en évidence, d'une part, des relations entre des objets et, d'autre part les relations entre ces objets et leurs paramètres. Il s'agit de construire une partition de l'ensemble des objets en un ensemble de classes qui soient les plus homogènes possible. La classification, a donc deux objectifs à atteindre :

- Trouver un modèle capable de représenter la répartition des données (catégorisation).
- Définir de manière formelle l'appartenance à l'une ou l'autre des classes de toute nouvelle donnée (généralisation).

En effet la classification a pour but de réduire l'espace de recherche dans une base de données lors du processus d'identification. En pratique, on peut rencontrer deux catégories de problèmes de classification, la classification supervisée et non-supervisée :

- Apprentissage supervisé:** dans ce type d'apprentissage, on cherche à : Estimer une fonction  $f(x)$  qui est la relation entre les objets et leurs classes. Les objets utilisés comme données d'apprentissages sont accompagnés par la classe à laquelle ils appartiennent.
- Apprentissage non supervisé:** on ne cherche pas cette fois à estimer une fonction mais on cherche à regrouper les objets ayant des caractéristique commune, les objets utilisés comme données d'apprentissage sont présentés sans leur classes.

### 4.1.3 Principe des machines à vecteurs de support :

Les SVM constituent une classe d'algorithmes basée sur le principe de minimisation de « Risque structurel » décrit par la théorie de l'apprentissage statistique qui utilise la séparation linéaire. Cela consiste à séparer par un hyperplan des individus représentés dans un espace de dimension égale au nombre de caractéristiques, les individus étant alors séparés en deux classes. Cela est possible quand les données à classer sont linéairement séparables. Dans le cas contraire, les données seront projetées sur un espace de plus grande dimension qu'elles deviennent linéairement séparables.

- **Linéarité et non-linéarité** : Parmi les modèles des SVM, on constate les cas linéairement séparables et les cas non linéairement séparable. Les premiers sont les plus simples de SVM car ils permettent de trouver facilement le classificateur linéaire. Dans la plupart des problèmes réels il n'y a pas de séparation linéaire possible entre les données, le classificateur de marge maximale ne peut pas être utilisé car il fonctionne seulement si les classes de données d'apprentissage sont linéairement séparables.

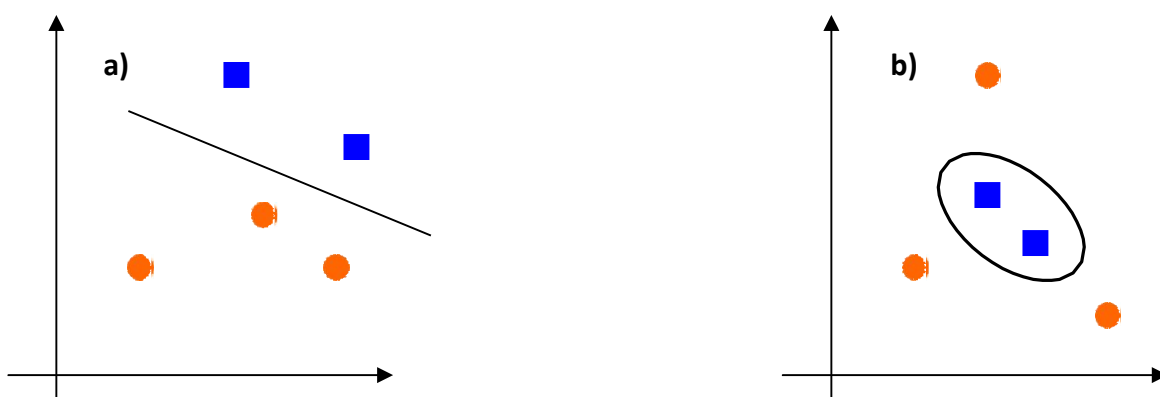


Figure 4.1 : a) Cas linéairement séparable, b) Cas non linéairement séparable

#### A) Cas de données linéairement séparables :

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classificateur est un classificateur linéaire appelé hyperplan.

Dans le schéma qui suit, on détermine un hyperplan qui sépare les deux ensembles de points.

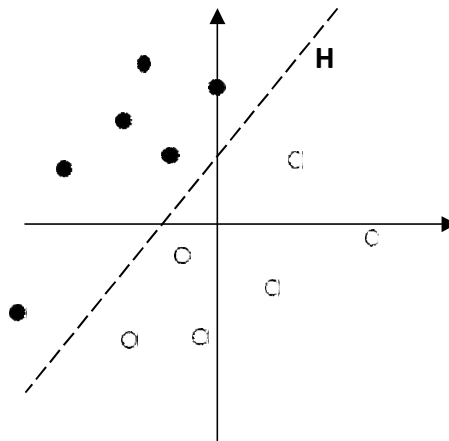


Figure 4.2 : Exemple d'un hyperplan séparateur

Les points les plus proches, qui seuls sont utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support.

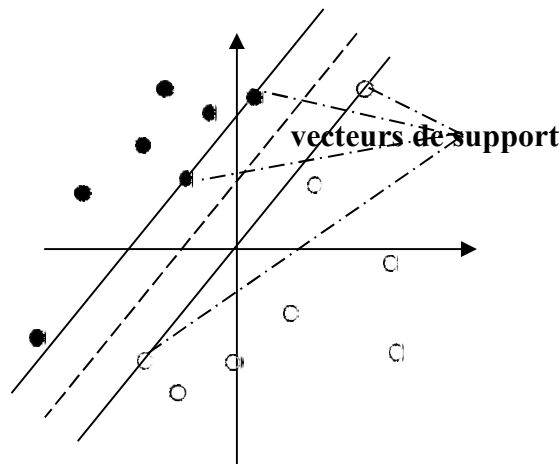


Figure 4.3 : Exemple de vecteurs de support

Il est évident qu'il existe une multitude d'hyperplan valide mais la propriété remarquable des SVM est que cet hyperplan doit être optimal (figure 4.4). Nous allons donc en plus chercher parmi les hyperplans valides, celui qui passe « au milieu » des points des deux classes d'exemples. Intuitivement, cela revient à chercher l'hyperplan le « plus sûr ».

En effet, supposons qu'un exemple n'ait pas été décrit parfaitement, une petite variation ne modifiera pas sa classification si sa distance à l'hyperplan est grande. Formellement, cela revient à chercher un hyperplan dont la distance minimale aux exemples d'apprentissage est maximale.

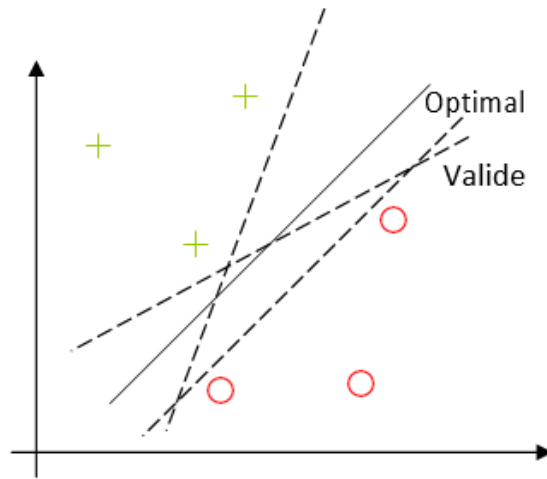


Figure 4.4 : Exemple de recherche d'un hyperplan optimal

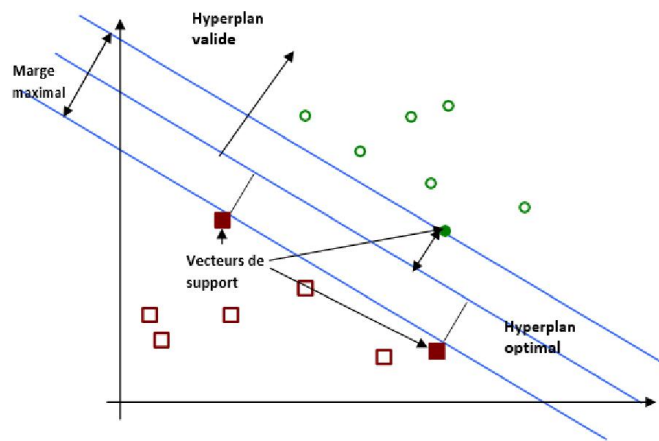


Figure 4.5 : Exemple de marge maximale (hyperplan valide).

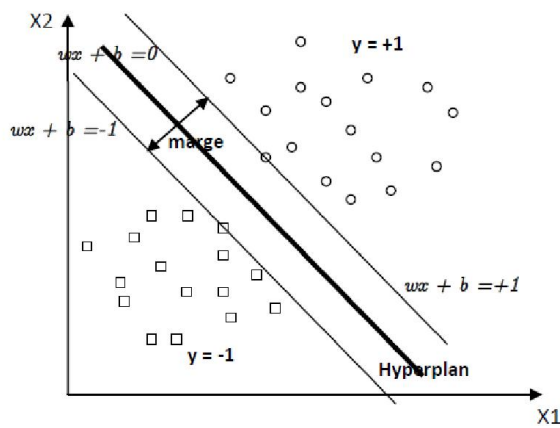


Figure 4.6 SVM binaire

- Pourquoi maximiser la marge ?

Intuitivement, le fait d'avoir une marge plus large procure plus de sécurité lorsque l'on classe un nouvel exemple. De plus, si l'on trouve le classificateur qui se comporte le mieux vis-à-vis des données d'apprentissage, il est clair qu'il sera aussi celui qui permettra au mieux de classer les nouveaux exemples. Dans le schéma qui suit, la partie droite nous montre qu'avec un hyperplan optimal, un nouvel exemple reste bien classé alors qu'il tombe dans la marge. On constate sur la partie gauche qu'avec une plus petite marge, l'exemple se voit mal classé.

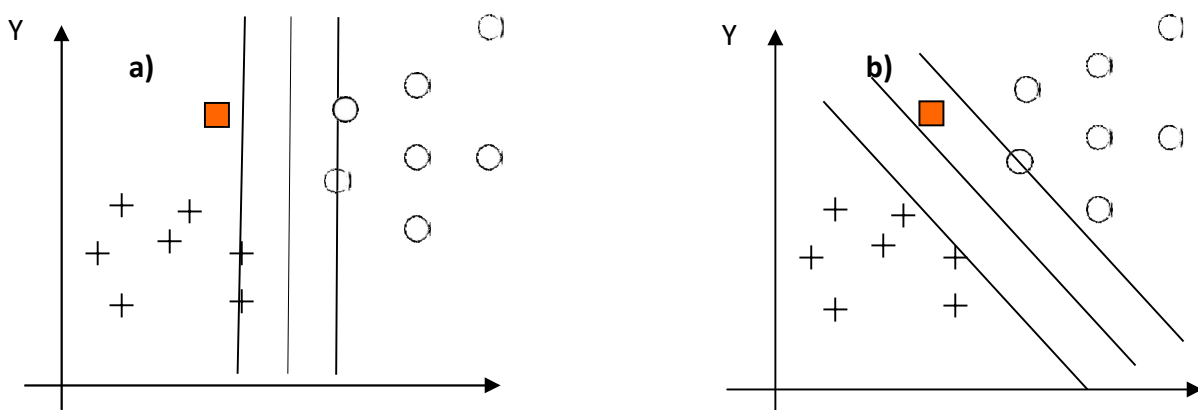


Figure 4.7 :a) Hyperplan avec faible marge, b) Meilleur hyperplan séparateur

En général, la classification d'un nouvel exemple inconnu est donnée par sa position par rapport à l'hyperplan optimal. Dans le schéma suivant, le nouvel élément sera classé dans la catégorie des « + ».

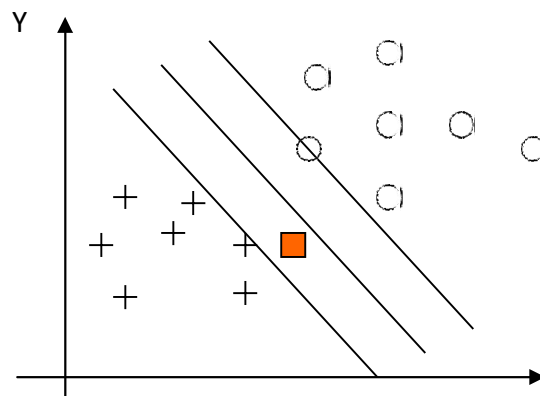


Figure 4.8 : Exemple de classification d'un nouvel élément.

- **Fondements mathématiques :**

Nous allons détailler dans les paragraphes ci-dessous les principes mathématiques sur lesquels repose SVM.

L'hyperplan séparateur est représenté par l'équation suivante :

$$H(x) = w^T x + b$$

Où  $w$  est un vecteur de  $m$  dimensions et  $b$  est un terme. La fonction de décision, pour un exemple  $x$ , peut être exprimé comme suit :

$$\begin{cases} \text{Classe} = 1 & \text{Si } H(x) > 0 \\ \text{Classe} = -1 & \text{Si } H(x) < 0 \end{cases}$$

Puisque les deux classes sont linéairement séparables, il n'existe aucun exemple qui se situe sur l'hyperplan, c.-à-d. qui satisfait  $H(x) = 0$ . Il convient alors d'utiliser la fonction de décisions suivante :

$$\begin{cases} \text{Classe} = 1 & \text{Si } H(x) > 1 \\ \text{Classe} = -1 & \text{Si } H(x) < -1 \end{cases}$$

Les valeurs  $+1$  et  $-1$  à droite des inégalités peuvent être des constantes quelconques  $+a$  et  $-a$ , mais en divisant les deux parties des inégalités par  $a$ , on trouve les inégalités précédentes qui sont équivalentes à l'équation :

$$y_i(w^T x_i + b) \geq 1; i = 1..n \quad (4.0)$$

L'hyperplan  $w^T x + b = 0$  représente un hyperplan séparateur des deux classes, et la distance entre cet hyperplan et l'exemple le plus proche s'appelle la marge (figure 4.5). La région qui se trouve entre les deux hyperplans  $w^T x + b = -1$  et  $w^T x + b = +1$  est appelée la région de généralisation de la machine d'apprentissage. Plus cette région est importante, plus est la capacité de généralisation de la machine.

La maximisation de cette région est l'objectif de la phase d'entraînement qui consiste, pour la méthode SVM, à rechercher l'hyperplan qui maximise la région de généralisation c.-à-d. la marge. Un tel hyperplan est appelé "hyperplan de séparation optimale". En supposant que les données d'apprentissage ne contiennent pas des données bruitées (mal-étiquetées) et que les données de test suivent la même probabilité que celle des données d'entraînement, l'hyperplan de marge maximale va certainement maximiser la capacité de généralisation de la machine d'apprentissage.

La détermination de l'hyperplan optimal passe par la détermination de la distance euclidienne minimale entre l'hyperplan et l'exemple le plus proche des deux classes. Puisque le vecteur  $w$  est orthogonal sur l'hyperplan séparateur, la droite parallèle à  $w$  et reliant un exemple  $x$  à l'hyperplan est donnée par la formule : (4.1)

$$\frac{aw}{\|w\|} + x = 0 \quad (4.1)$$

Où  $a$  représente la distance entre  $x$  et l'hyperplan. La résolution de (4.1), donne (4.2) :

$$a = -\frac{w^T x + b}{\|w\|} \quad (4.2)$$

La distance de tout exemple de l'hyperplan doit être supérieure ou égale à la marge  $\delta$

$$\frac{y_i(w^T x_i + b)}{\|w\|} \geq \delta \quad (4.3)$$

Si une paire  $(w, b)$  est une solution alors  $(a w; a b)$  est une solution aussi où  $a$  est un scalaire. On impose alors la contrainte (4.4) suivante :

$$\|w\| \delta \geq 1 \quad (4.4)$$

Pour trouver l'hyperplan séparateur qui maximise la marge, on doit déterminer, à partir des deux dernières inégalités, le vecteur  $w$  qui possède la norme euclidienne minimale et qui vérifie la contrainte de l'équation (4.0), de bonne classification des exemples d'entraînement. L'hyperplan séparateur optimal peut être obtenu en résolvant le problème de l'équation (4.5).

$$\begin{cases} \text{Minimiser} & \frac{1}{2} \|w\|^2 \\ \text{sous contraintes} & \\ y_i(w^T x_i + b) \geq 1 & \forall i = 1..n \end{cases} \quad (4.5)$$

Remarquons que nous pouvons obtenir le même hyperplan même en supprimant toutes les données qui vérifient l'inégalité de la contrainte. Les données qui vérifient l'égalité de la contrainte s'appellent les vecteurs supports, et ce sont ces données seules qui contribuent à la détermination de l'hyperplan. Dans la figure (4.6), les données qui se trouvent sur les deux droites  $+1$  et  $-1$  représentent les vecteurs supports.

Le problème de l'équation (4.5) est un problème de programmation quadratique avec contraintes linéaires. Dans ce problème, les variables sont  $w$  et  $b$ , c.-à-d. que le nombre de variables est égal à  $m + 1$ .

Généralement, le nombre de variables est important ce qui ne permet pas d'utiliser les techniques classiques de programmation quadratique. Dans ce cas le problème (4.5) est convertit en un problème dual équivalent sans contraintes de l'équation (4.6) qui introduit les multiplicateurs de Lagrange :

$$Q(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^n \alpha_i \{y_i(w^T x_i + b) - 1\} \quad (4.6)$$

Où les  $\alpha_i$  sont les multiplicateurs non négatifs de Lagrange. L'optimum de la fonction objective  $Q$  peut être obtenu en la minimisant par rapport à  $w$  et  $b$  et en la maximisant par rapport aux  $\alpha_i$ . À l'optimum de la fonction objective, ses dérivées par rapports aux variables  $w$  et  $b$  s'annulent ainsi que le produit des  $\alpha_i$  aux contraintes (équation 4.7) :

$$\begin{cases} \frac{\partial Q(w,b,\alpha)}{\partial w} = 0 & (a) \\ \frac{\partial Q(w,b,\alpha)}{\partial b} = 0 & (b) \\ \alpha_i \{y_i(w^T x_i + b) - 1\} = 0 & (c) \\ \alpha_i \geq 0 & (d) \end{cases} \quad (4.7)$$

De (4.7.a) on déduit :

$$\begin{cases} w = \sum_{i=1}^n \alpha_i y_i x_i \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (4.8)$$

En remplaçant dans 4.6, on obtient le problème dual à maximiser suivant :

$$\begin{cases} \text{Maximiser} & Q(\alpha) = \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{Sous contraintes} & \sum_{i=1}^n \alpha_i y_i = 0 \\ & \alpha_i \geq 0 \end{cases} \quad (4.9)$$

Si le problème de classification est linéairement séparable, une solution optimale pour les  $\alpha_i$  existe. Les exemples ayant des  $\alpha_i \neq 0$  représentent les vecteurs supports appartenant aux deux classes.

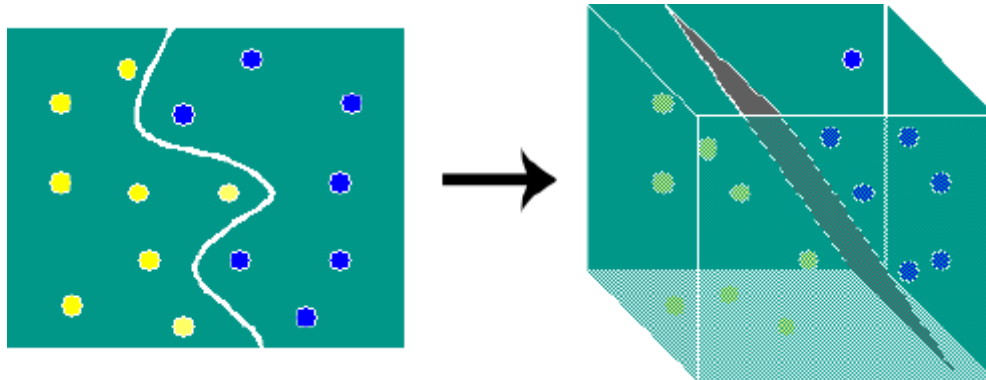
La fonction de décision est donnée par :

$$H(x) = \sum_S \alpha_i y_i x^T x_i + b \quad (4.10)$$

Où  $S$  représente l'ensemble des vecteurs supports.

### B) Cas de données non linéairement séparables :

Pour surmonter les inconvénients des cas non linéairement séparables, l'idée des SVM est de changer l'espace des données. La transformation non linéaire des données peut permettre une séparation linéaire des exemples dans un nouvel espace. On va donc avoir un changement de dimension. Cette nouvelle dimension est appelé « espace de ré-description ». En effet, intuitivement, plus la dimension de l'espace de ré-description est grande, plus la probabilité de pouvoir trouver un hyperplan séparateur entre les exemples est élevée. Ceci est illustré par le schéma suivant:



**Figure 4.9:** Exemple de changement de l'espace de données

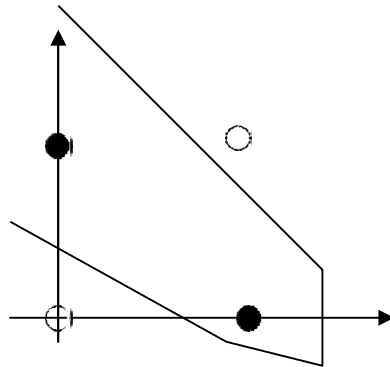
On a donc une transformation d'un problème de séparation non linéaire dans l'espace de représentation en un problème de séparation linéaire dans un espace de ré-description de plus grande dimension. Cette transformation non linéaire est réalisée via une fonction noyau.

En pratique, quelques familles de fonctions noyau paramétrables sont connues et il revient à l'utilisateur de SVM d'effectuer des tests pour déterminer celle qui convient le mieux pour son application. On peut citer les exemples de noyaux suivants : polynomiale, gaussien, sigmoïde et Laplacien.

- **Illustration de transformation de cas non linéaire : le cas XOR :**

Le cas de XOR n'est pas linéairement séparable, si on place les points dans un plan à deux dimensions, on obtient la figure suivante :

Coordonnées des points : (0,0) ; (0,1) ; (1,0) ; (1,1)



**Figure 4.10 :** Illustration de cas non linéairement séparable (le cas XOR).

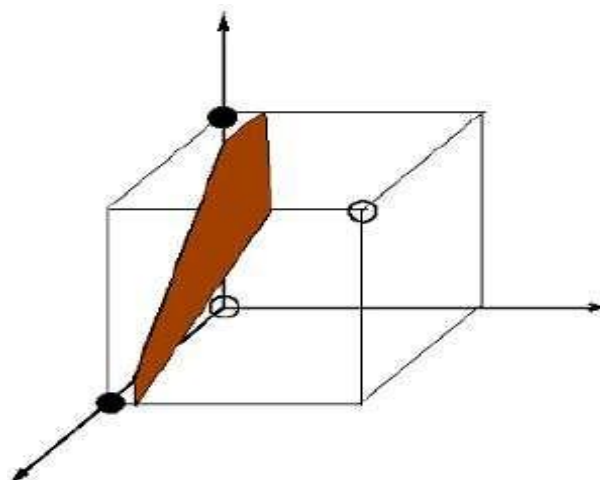
Si on prend une fonction polynomiale  $(x,y) \rightarrow (x,y,x.y)$  qui fait passer d'un espace de dimension 2 à un espace de dimension 3, on obtient un problème en trois dimensions linéairement séparable :

$$(0,0) \rightarrow (0,0,0)$$

$$(0,1) \rightarrow (0,1,0)$$

$$(1,0) \rightarrow (1,0,0)$$

$$(1,1) \rightarrow (1,1,1)$$



**Figure 4.11 :** Illustration de passage d'un espace 2D à un espace 3D .

Alors Pour le cas non linéaire, on doit suivre les mêmes étapes que dans le cas linéaire. Néanmoins, un changement d'espace pour les données  $x_i$  est nécessaire. Cela est dû à la projection des données dans un espace de dimension supérieur qui rend le problème complètement linéaire.

La projection est faite par une fonction  $k(x_i, x_j)$  qui est appelé fonction de noyau.

Les noyaux les plus utilisés sont définis comme suit :

- **Noyau Linéaire** :  $k(x_i, x_j) = \langle x_i, x_j \rangle$  ;
  - **Noyau Polynomial** :  $k(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + r)^d, \gamma > 0$  ;
  - **Noyau à base Radial (RBF)** :  $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$  ;
  - **Noyau Sigmoidé** :  $k(x_i, x_j) = \tanh(\gamma \langle x_i, x_j \rangle + r)$  ;
- où  $\gamma, r$  et  $d$  sont les paramètres du noyau.

#### 4.1.4 SVM pour la régression(SVR) :

Dans leur origine, les SVMs ont été développées pour des problèmes de classification. Cependant, leur nature leur permet de résoudre également des problèmes de régression. La régression est un cas particulier de classification où les classes des exemples ne sont pas dénombrables c.-à-d. continues. Le problème consiste à trouver, en utilisant  $D = \{(x_1, y_1), \dots, (x_n, y_n)\}$  une fonction  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  qui rapproche le plus possible des  $y_i$ , en d'autre terme qui minimise la différence entre les  $f(x_i)$  et les  $y_i$ .

Souvent,  $f$  est considérée comme fonction linéaire :  $f = \langle w, x \rangle + b$  où  $w$  est un vecteur et  $b$  est un scalaire.

Le problème revient donc à trouver un hyperplan caractérisé par  $w^*$  et  $b^*$  qui minimise l'écart global entre  $f$  et les  $y_i$ .

Pour résoudre ce problème, les SVMs utilisent une astuce semblable a celle utilisée en classification. On propose de modéliser la fonction de régression par un hyperplan qui se situe au centre d'un hyper-tube de largeur  $2\epsilon$  contenant tous les exemples d'entraînement.

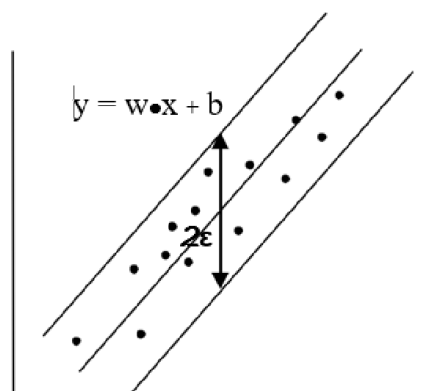


Figure 4.12 – Hyper-tube modélisant la fonction de régression

Plusieurs hyper-tubes, de largeur  $2\epsilon$  contenant tous les exemples d'entraînement, peuvent exister. L'hyper-tube optimal est celui qui minimise la distance entre les exemples d'entraînement et ses frontières, autrement dit, qui maximise la distance des exemples de l'hyperplan du centre. La détermination de l'hyper-tube optimal est semblable à la détermination de l'hyperplan optimal de marge maximale dans le cas de classification. On doit donc rechercher un hyper-tube de marge maximale avec tous les exemples d'entraînement à l'intérieur. Par une analyse similaire à celle du problème de classification, la solution du problème de régression est réduite à la résolution du problème dual d'optimisation quadratique de l'équation :

$$\left\{ \begin{array}{l} \text{Maximiser } \alpha, \alpha' \\ \quad -1/2 \sum_{i,j=1}^n (\alpha_i - \alpha_j') \langle x_i, x_j \rangle \\ \quad - \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i') + \sum_{i=1}^n y_i (\alpha_i - \alpha_i') \\ \text{sous contraintes} \\ \quad \sum_{i=1}^n (\alpha_i - \alpha_i') = 0 \\ \quad 0 \leq \alpha_i - \alpha_i' \leq C \end{array} \right.$$

Où les  $\alpha_i$  et les  $\alpha_i'$  sont les coefficients des exemples respectivement au-dessus et au-dessous de l'hyperplan et  $C$  est un paramètre pour leur pénalisation. La fonction de sortie  $f(x)$  peut être donnée par l'équation :

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i') \langle x_i, x \rangle + b$$

## 4.2 Différents modèles de combinaison binoculaire:

Pour la perception d'une image 3D une opération de combinaison de deux vues gauche et droite doit être faite et pour une meilleure compréhension des études physiologiques de la vision binoculaire est indispensable pour le développement de nouvelles métriques. Pour cela, différents modèles de combinaison binoculaires ont été proposés dans la littérature. Parmi ces modèles, on a la combinaison binoculaire de la luminosité ou du contraste.

La perception binoculaire de luminosité pour une image stéréoscopique peut être donnée par la fonction  $f_B$  de la luminosité des vues gauche et droite comme suit :

$$I_{Binoc} = f_B(I_G, I_D) \quad (4.13)$$

Avec  $I_{Binoc}$  l'image perçue après combinaison et  $I_G, I_D$  les deux vues gauche-droite respectivement. La fonction de combinaison  $f_B$  est exprimée à l'aide de différents modèles, pondération des vues, sommation quadratique et sommation de vecteur.

- **Modèle de pondération des vues** : ce modèle décrit la combinaison binoculaire comme la somme pondérée de l'image gauche et droite. La fonction  $f_B$  est donnée par l'équation (4.12) suivante :

$$f_B = w_1 \cdot I_G + w_2 \cdot I_D \quad (4.14)$$

avec  $w_1, w_2$  sont les poids de l'image gauche et droite respectivement.

- **Modèle de sommation quadratique** : pour ce modèle la combinaison binoculaire est exprimée par la sommation quadratique des deux luminances gauche-droite comme suit :

$$f_B = \sqrt{I_G^2 + I_D^2} \quad (4.15)$$

Ce modèle est valable pour des luminances symétriques pour les deux yeux.

- **Modèle de sommation de vecteur** : ce modèle est basé sur les signaux de flux de luminosité monoculaire de l'œil gauche et de l'œil droit comme suit :

$$f_B = \frac{\epsilon_G^2 + \epsilon_D^2}{\epsilon_G + \epsilon_D} \quad (4.16)$$

avec  $(\epsilon_G, \epsilon_D)$  les flux de signaux des deux images gauche et droite.

- **Modèle des réseaux de neurones** : il utilise les réponses neuronales de chaque œil pour le calcul de la combinaison binoculaire.

$$f_B = N_G + N_D \quad (4.17)$$

avec  $N_G, N_D$  les réponses neuronales des yeux gauche et droit respectivement.

- **Modèle de théorie de gain-contrôle** : les auteurs ont proposé une étude pour la combinaison binoculaire basée sur deux principes d'interaction. Pour chaque voisinage spatial, chaque œil exerce un contrôle de gain sur le signal de l'autre œil en proportion de l'énergie de contraste. Ce dernier exerce aussi un contrôle de gain sur celui de l'autre œil. En tenant compte de ces spécificités, un simple modèle physiologique qui décrit la combinaison binoculaire a été proposé comme suit :

$$\begin{aligned} f_B &= \left( \frac{1 + Eng_G}{1 + Eng_G + Eng_D} \right) I_G + \left( \frac{1 + Eng_D}{1 + Eng_G + Eng_D} \right) I_D \\ &= G_G \cdot I_G + G_D \cdot I_D \end{aligned} \quad (4.18)$$

### 4.3 Carte de disparité:

#### 4.3.1 définition de disparité :

Dans le domaine de la vision stéréoscopique le terme « *disparité* » décrit la différence entre les coordonnées des points (pixels) homologues sur les plans de la paire d'images stéréoscopique. La valeur de la disparité est inversement proportionnelle à la profondeur. Par conséquent, des grandes valeurs de la disparité correspondent aux points proches des caméras. Autrement dit, plus un point de la scène est loin de la capture, plus la disparité entre ses points homologues est réduite, et vice versa.

Etant donné, un point P de l'espace dont  $p(x_l, y_l)$  et  $p'(x_r, y_r)$  et ses projections perspectives sur les deux plans images gauche et droite, respectivement. La disparité du point P peut être définie comme étant une fonction  $d_l$  permettant l'attribution une valeur de disparité à chaque pixel de l'image gauche,

$$\begin{aligned} d_l: \quad \mathbb{N}^2 &\rightarrow \mathbb{R}^2 \\ (x_l, y_l) &\rightarrow d_l(x_l, y_l) = (x_l - x_r, y_l - y_r) \end{aligned} \quad (4.19)$$

Dans une configuration parallèle où les images gauche et droite sont rectifiées, les points homologues P et P' se trouvent sur les lignes de même ordonnées de l'image gauche et droite respectivement, c'est-à-dire  $y_l = y_r$ . La fonction (4.17) peut être simplifiée comme suit :

$$\begin{aligned} d_l: \quad \mathbb{N}^2 &\rightarrow \mathbb{R} \\ (x_l, y_l) &\rightarrow d_l(x_l, y_l) = x_l - x_r \end{aligned} \quad (4.20)$$

Lorsque l'image droite est considérée comme image de référence, la disparité à chaque pixel

$(x_r, y_r)$  de l'image droite s'écrit comme suit :

$$\begin{aligned} d_r: \quad \mathbb{N}^2 &\rightarrow \mathbb{R} \\ (x_r, y_r) &\rightarrow d_r(x_r, y_r) = x_l - x_r \end{aligned}$$

A partir des équations (4.18)et (4.19), nous pouvons déduire pour tout point  $(x_l, y_l)$  de l'image gauche, l'égalité suivante :

$$d_l(x_l, y_l) = d_r(x_l - d_l(x_l, y_l), y_l) \quad (4.21)$$

Le processus de la recherche de la disparité pour toutes les primitives extraites de l'image dans une image appelée *carte de disparité*, cette dernière est une image dont la valeur de chaque pixel

correspond à la valeur de la disparité associée. La carte de disparité est utilisée afin de simplifier la visualisation des résultats de la mise en correspondance : plus le pixel est clair plus la disparité est importante, et vice versa. Comme il montré dans la figure 4.22.

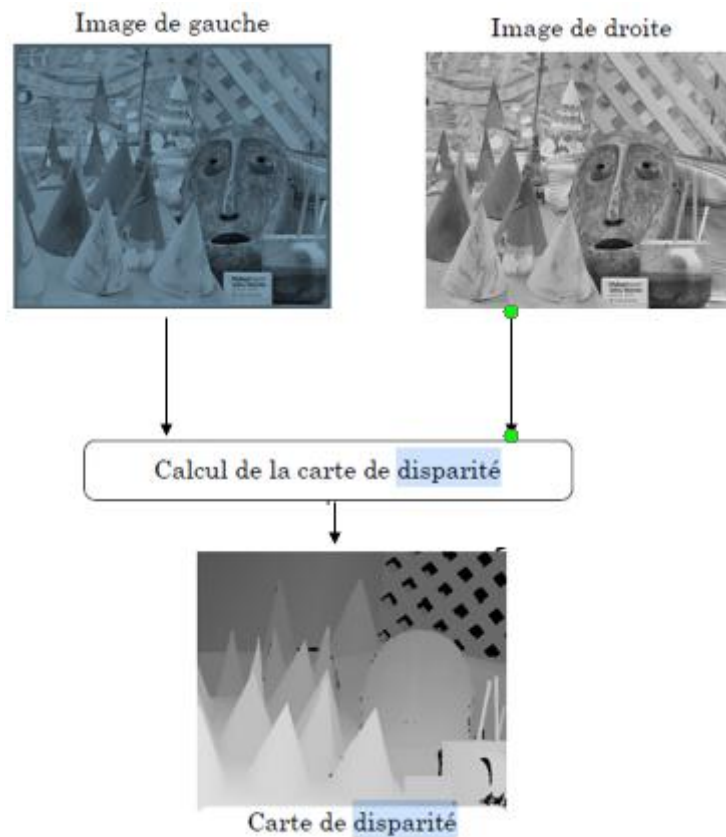


Figure 4.22 l'image de disparité

#### 4.4 Modèle d'activité spatiale binoculaire :

Afin d'analyser les mécanismes de la vision binoculaire d'une manière accessible, nous étudions en profondeur les phénomènes de fusion binoculaire, de rivalité et de suppression, La relation entre la quantité d'activité spatiale, qui est une mesure de la variance d'image normalisée, contenue dans une paire d'images S3D et la qualité d'image perçue est également étudiée, Levelt a démontré que les facteurs sensoriels de bas niveau peuvent fortement influencer la rivalité et la suppression binoculaires.

Il a souligné que des stimuli puissants, tels que des zones de contraste élevé ou des bords nets tendent à dominer la rivalité entre les deux vues. En se basant sur cette idée, nous construisons un modèle de fusion cyclopéen basé sur l'hypothèse qu'une image dans une paire stéréoscopique (en niveaux de gris) contenant la plus grande quantité de l'activité spatiale dominera la perception.

#### **4.4.1 Fusion binoculaire, rivalité et suppression :**

Le système visuel humain (SVH) reçoit des stimuli visuels du deux yeux et les combine en un seul percept combiné. Lorsque les deux vues partagent des caractéristiques similaires, la fusion binoculaire peut se déroulent facilement, grâce à des processus de vergence, d'accommodation et d'autres adaptations visuelles lors de la visualisation du monde réel et naturel. Cependant, une rivalité binoculaire peut survenir lorsque des stimuli monoculaires inhabituels différents (dichoptiques) sont présentés aux emplacements rétiniens correspondants des deux yeux. Ensuite, plutôt que de percevoir une seule image fusionnée stable, on expérimente soit des alternances de conscience perceptuelle au fil du temps, alors que les deux stimuli se disputent la domination perceptive, ce qui se produit généralement lors de la visualisation de stéréopaires contenant des distorsions asymétriques.

Si l'un des stimuli domine l'autre, alors la suppression binoculaire se produira. Une question difficile connexe est de savoir comment modéliser le processus de combinaison binoculaire, lors de l'évaluation de stimuli asymétriquement dégradés. Étant donné que la perception binoculaire des degrés ou des caractéristiques de distorsion peut grandement influencer la qualité perçue des images S3D, déterminer comment synthétiser les deux vues pour simuler la perception d'une scène 3D est un élément principale d'un modèle d'évaluation de la qualité des images 3D réussi.

La visualisation normale d'images stéréo dans la vie quotidienne se traduit généralement par une fusion binoculaire plutôt qu'une rivalité binoculaire, car la stéréopaire partage un contenu similaire. Cependant, lors de la visualisation du contenu S3D affiché électroniquement, l'introduction de distorsions peut détruire cette similitude, conduisant à une rivalité binoculaire. Cette rivalité peut modifier la perception des distorsions, et peut même conduire à des sensations d'inconfort physiologique. De toute évidence, la rivalité binoculaire est un facteur important dans la perception de la qualité S3D.

#### **4.4.2 Information spatiale et perception binoculaire :**

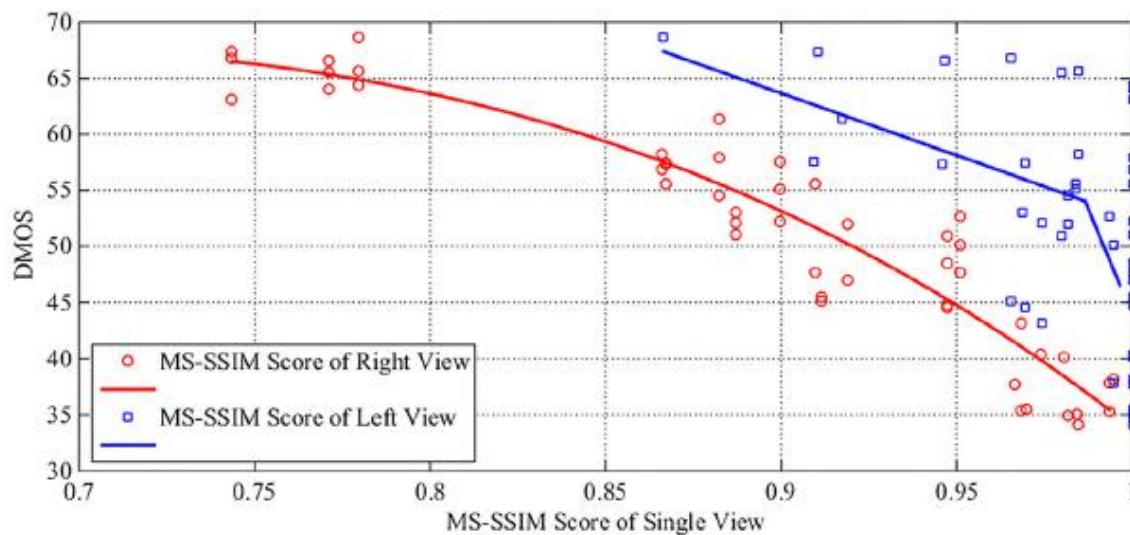
Nous avons mené des études sur le flou et le bruit des images dégradées pour clarifier davantage le lien entre la vision binoculaire et ces distorsions. Si l'une des vues est floue et l'autre nette, le système visuel peut réduire la perte d'informations à haute fréquence de la vue floue, ce qui entraîne une perception stéréoscopique fusionnée très semblable à la vue plus nette. La perception des stéréo paires a tendance à être dominée par l'image plus nette et de meilleure qualité lorsque l'une ou les deux sont dégradées par le flou. Cependant, lorsque du bruit est introduit, la perception des stéréopaires peut être dominée par le composant de faible qualité, c'est-à-dire une distorsion à haute fréquence.

Le flou a tendance à diminuer la quantité d'activité spatiale dans les images, tandis que le bruit a tendance à augmenter la quantité d'activité spatiale, c'est-à-dire qu'en comparant la quantité d'activité spatiale dans les vues gauche et droite, nous pouvons déterminer les facteurs dominants affectant la

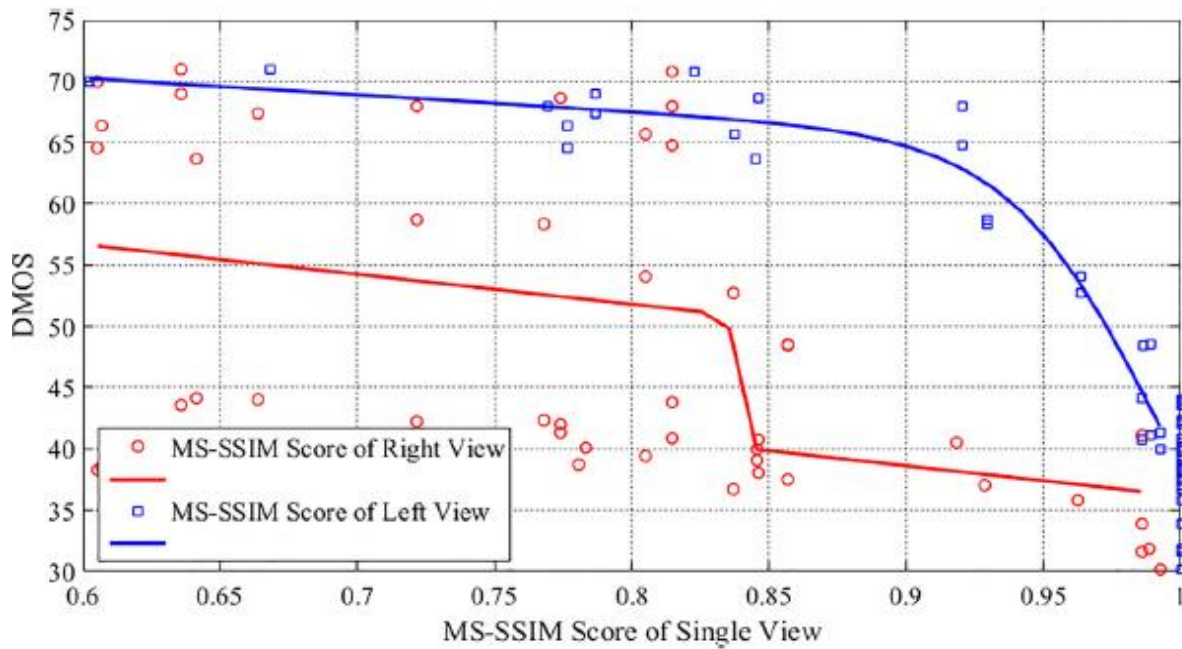
prédiction de la qualité S3D. Ainsi, notre approche de la prédiction de la qualité S3D fonctionne sous l'hypothèse que la perception de la qualité S3D a tendance à être dominée par l'image contenant la plus grande quantité d'activité spatiale.

Pour valider cette hypothèse, nous avons effectué des séries d'expériences sur un sous-ensemble de la base de données LIVE 3D Image Quality. Plus précisément, nous avons utilisé l'ensemble des stéréopaires asymétriques dégradées par le flou et le bruit. De par leur conception, toutes les vues de gauche sélectionnées ont moins de distorsion que les vues de droite correspondantes dans la base de données LIVE. Nous avons utilisé l'indice MS-SSIM comme métrique pour mesurer la qualité de chacune des images gauche et droite.

Nous avons utilisé l'image individuelle (gauche et droite) de la référence et des paires stéréoscopiques dégradées comme entrées pour calculer les scores MS-SSIM, puis analysé la corrélation entre les scores MS-SSIM et le DMOS fourni par la base de données (voir figures 4.23 et 4.24). Pour le cas de la distorsion du bruit, les résultats montrent que la vue de droite (qui contient plus d'informations spatiales que la vue de gauche) les scores MS-SSIM sont mieux corrélés avec les jugements humains (DMOS) que les scores de la vue de gauche. Pour la distorsion de flou, la vue de gauche (avec plus d'informations spatiales que la vue de droite) MS-SSIM est mieux corrélée avec la perception de la qualité humaine. Ces résultats expérimentaux soutiennent notre hypothèse, suggérant que la vue (gauche ou droite) contenant une plus grande activité spatiale dominera la perception de la qualité S3D.



**Figure. 4.23.** Nuage de points de DMOS par rapport aux scores MS-SSIM correspondants des vues gauche / droite asymétriquement dégradées par le bruit. Les courbes ajustées montrent que DMOS est en corrélation avec les scores MS-SSIM des images de vue de droite mieux qu'avec les images de vue de gauche. Le SROCC entre les scores DMOS et MS-SSIM de bonnes vues était de 0,9369, tandis que le SROCC entre DMOS et MS-SSIM



**Figure. 4.24.** Nuage de points de DMOS par rapport aux scores MS-SSIM correspondants des vues gauche / droite asymétriquement dégradées par le flou, Les courbes ajustées montrent que DMOS est en corrélation avec les scores MS-SSIM des images de vue de gauche mieux qu'avec les images de vue de droite, Le SROCC entre les scores DMOS et MS-SSIM de bonnes vues était de 0,5564, tandis que le SROCC entre DMOS et MS-SSIM le score des vues de gauche était de 0,8729.

#### 4.4.3 Modèle d'activité spatiale binoculaire :

À partir des expériences que nous avons décrites ci-dessus, nous avons observé que l'activité spatiale n'est pas seulement utile pour déterminer s'il y aura une rivalité binoculaire (suppression) entre les images de vue gauche et droite, mais elle peut également être utilisée pour évaluer davantage les degrés relatifs d'influence des deux vues lorsque la rivalité binoculaire se produit. Nous décrivons ensuite le modèle d'activité spatiale binoculaire quantitative exprimant notre hypothèse. La stéréoscopie est en grande partie un processus local, et les quantités relatives d'informations contenues dans les stéréopaires peuvent varier localement dans l'espace, c'est pourquoi nous définissons une mesure d'activité binoculaire spatiale locale comme suit. Soit  $I_L(i)$  l'image de la vue de gauche (luminance), où  $i = (x, y)$  sont les coordonnées spatiales de l'image. Alors, soit  $\mathbf{S}_L(i)$  un voisinage centré sur  $i$ , par ex. défini par une fenêtre mobile  $N \times N$ . Nous avons observé que la performance du modèle n'est pas très sensible à la taille de la fenêtre, et nous l'avons fixé à  $N = 17$ . Puis exprimez l'activité spatiale dans  $\mathbf{S}_L(i)$  comme:

$$\epsilon [\mathbf{S}_L(i)] = \log_2 [\sigma_L^2(i) + 1], \quad (1)$$

où la constante additive unitaire garantit une information spatiale positive, et  $\sigma_L^2(i)$  est la variance de  $\mathbf{S}_L(i)$ . Le logarithme monotone sert à compresser les valeurs d'activité dans une plage plus petite.

De même définir les grandeurs  $\mathbf{S}_R(i), \sigma_R^2(i)$  et  $[\mathbf{S}_R(i)]$  sur l'image de droite  $I_R(i)$ , une image d'activité

spatiale «cyclopéenne» synthétisée  $O(i)$  peut être modélisée comme:

$$O(i) = \frac{\{\epsilon [S_L(i)] + C\} \cdot I_L(i) + \{\epsilon [S_R(i + d_i)] + C\} \cdot I_R(i + d_i)}{\{\epsilon [S_L(i)] + C\} + \{\epsilon [S_R(i + d_i)] + C\}} \quad (2)$$

$$P(i) = I_L(i) \cdot I_R(i + d_i) \quad (3)$$

Où  $d_i$  est la disparité horizontale à la coordonnée  $i$ , et  $C = 0,01$  est un petit nombre positif pour garantir la stabilité. Ainsi, le calcul de la mesure d'activité spatiale binoculaire  $O(i)$  suit trois étapes: (1) Générer une carte de disparité entre les images de vue gauche et droite à l'aide d'un algorithme d'estimation de disparité stéréo simple (nous déployons un algorithme stéréo simple qui utilise le SSIM index comme critère d'appariement), (2) Générer des cartes d'activité spatiale sur les images de vue gauche et droite en utilisant (1), (3) Synthétiser une image finale de l'activité spatiale «cyclopéenne» à l'aide de (2).

#### 4.4.4 Évaluation de la qualité des stéréopaires sans référence :

Le diagramme suivant montre le modèle de l'évaluation sans référence de la qualité d'une image S3D, qui nous permet de prédire la qualité de l'image stéréoscopique 3D, où nous avons une image droite et une autre gauche.

En premier lieu l'activité spéciale binoculaire est calculé, et calculer aussi l'image de produit, ensuite il va être extrait de l'images les caractéristique d'entropies spatiale et spectrale.

De ces caractéristiques on extrait un modèle d'évaluation de la qualité d'image des S3D par le SVR .

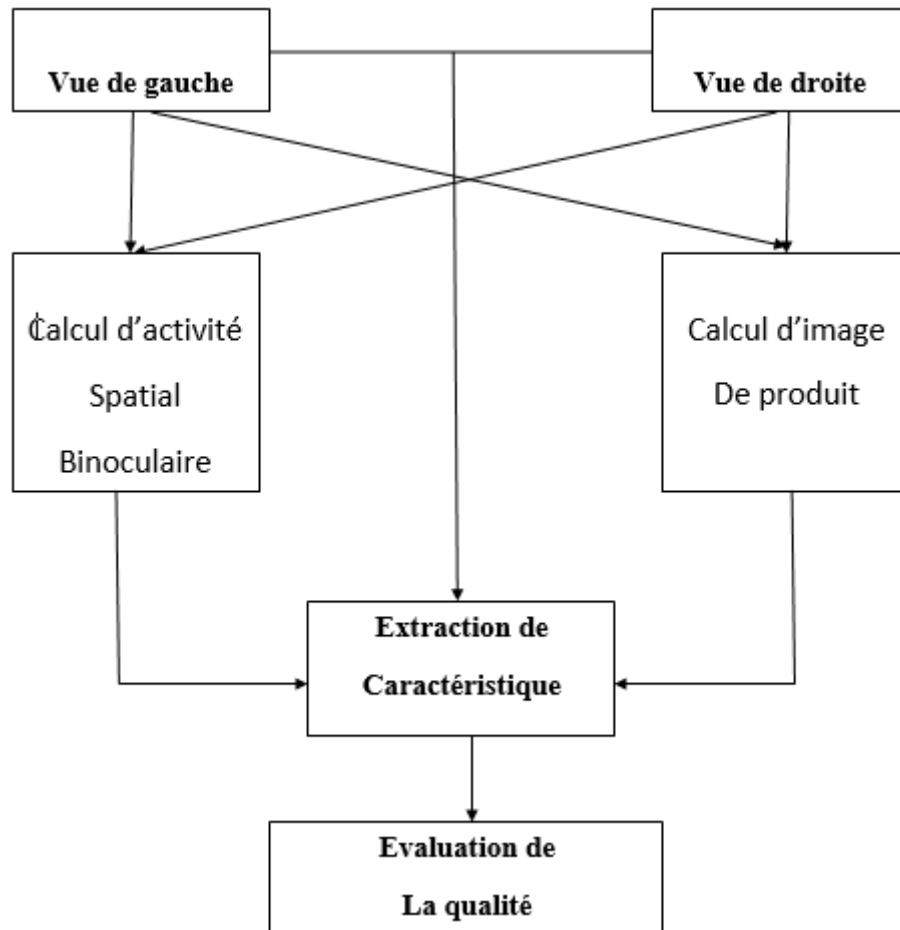


Figure 4.25 diagramme du modèle proposé d'évaluation de la qualité NR(sans référence) des images S3D

#### 4.5 Caractéristiques d'entropie d'image:

L'entropie de l'image indique la quantité d'informations contenue dans une image, et lorsqu'elle est calculée sur multi-échelle, il révèle l'entropie statistique de l'espace d'échelle. Les types et degrés de distorsions d'image affectent généralement l'entropie locale des images de manières diverses mais prévisibles. Contrairement à l'entropie locale, l'entropie globale de l'image capture les informations globales des images mais ne pas discriminer la distribution spatiale des informations. Ainsi, les images ayant la même entropie globale peuvent apparaître très différentes.

Des travaux antérieurs ont montré qu'il existe une relation étroite entre l'entropie locale de l'image et la qualité d'image perçue. De même, nous avons trouvé que les caractéristiques d'entropie sont très sensibles aux degrés et aux types de distorsions d'image. Notre méthode utilise des entropies calculées à partir de blocs d'image locaux, sur les réponses de l'échelle spatiale de bloc et les DCT coefficients du bloc. Toutes les entropies sont calculées localement. L'entropie spatiale est fonction de la probabilité de

distribution du pixel local, tandis que l'entropie spectrale est fonction de la distribution de probabilité des valeurs locales de coefficient DCT.

Par rapport aux méthodes IQA basées sur NSS, qui utilisent les caractéristiques statistiques au niveau des pixels, notre méthode basée sur l'entropie analyse la distribution conjointe des pixels dans un patch local. Notre méthode reflète les caractéristiques statistiques des régions locales plutôt que des pixels, qui contribueront à décrire l'information structurale locale des images.

Notre hypothèse de base est que l'entropie locale des images non dégradés possède certaines propriétés statistiques. Les propriétés statistiques résultent de la dépendance entre les pixels adjacents. Et l'introduction de la distorsion détruira cette dépendance inhérente, produisant changement d'entropie locale.

Par exemple, un bruit blanc additif entraînera des valeurs d'entropie locale plus élevées en raison de l'introduction de beaucoup d'informations à haute fréquence sur des images non dégradées, mais une distorsion de flou réduira les valeurs d'entropie locale en raison de la perte de détails de l'image.

Ici, nous allons décrire l'entropie locale et discuter en détail de la relation entre l'entropie locale et la qualité de l'image.

#### 4.5.1 Caractéristiques de l'entropie spatiale : (f1-f6)

L'entropie spatiale est : 
$$E_s = - \sum_x p(x) \log_2 p(x),$$

Où  $x$  sont les valeurs de pixel dans un bloc, avec probabilité densité empirique  $p(x)$ , c'est-à-dire fréquence relative. Pour illustrer le comportement des valeurs d'entropie spatiale locale contre différents degrés et types de distorsions, nous avons mené une série d'expériences de validation sur une image. Comme le montre la figure 4.26, différents types de distorsions (compression JP2 et JPEG, bruit, flou et décoloration rapide) exercent systématiquement des influences différentes sur les valeurs de l'entropie spatiale.

L'image non dégradé (ori) a une histogramme d'entropie spatiale avec une moyenne d'environ 4 et qui est asymétrique à gauche. Le "biais à gauche" signifie qu'il y a moins de données à gauche de valeur moyenne que droite, conduisant à une queue gauche plus longue que droite. Cependant, l'introduction de la distorsion modifie sa moyenne et son biais. Par exemple, le bruit augmente fortement la moyenne, tandis que flou et jp2k réduisaient la moyenne et inclinent l'histogramme vers la droite. L'histogramme jpeg est assez différent car de nombreux zéros se produisent en diminuant fortement l'entropie spatiale locale. Les caractéristiques d'entropie spatiale sont fortement indicatives du type de distorsion comme nous le verrons ensuite.

Nous utilisons la moyenne et le biais (inclination) comme caractéristiques de qualité qui décrivent l'histogramme. Nous avons extrait ces 2 caractéristiques de chaque échelle, ce qui donne  $2 \times 3 = 6$  fonctionnalités.

#### 4.5.2 Caractéristiques de l'entropie spectrale :(f7-f12)

Puisque nous pensons qu'il existe une forte relation entre les valeurs d'entropie spectrale et le degré et le type de distorsion, la matrice C de coefficient DCT de bloc est également calculée sur  $8 \times 8$  blocs.

Nous normalisons les coefficients DCT pour la production carte de probabilité spectrale:

$$P(i,j) = \frac{C(i,j)^2}{\sum_i \sum_j C(i,j)^2}$$

Où  $1 \leq i \leq 8$ , et  $i, j \neq 1$  (DC est exclu).

Ensuite définir l'entropie spectrale locale :  $E_f = - \sum_i \sum_j P(i,j) \log_2 P(i,j)$ .

Pour explorer le comportement des caractéristiques d'entropie spectrale contre différents types de distorsions, nous avons mené une autre expérience de validation visuelle sur les mêmes images de la figure 4.23. Comme le montre la figure 4.24, différents types de distorsions (compression jp2k et jpeg, bruit, flou et décoloration rapide) exercent systématiquement des influences différentes sur les valeurs d'entropie spectrale. De toute évidence, nous pouvons constater que l'image non dégradé (ori) a un histogramme d'entropie spectrale qui est généralement incliné vers la gauche.

Cependant, l'introduction de la distorsion changera sa moyenne et son biais. Par exemple, le bruit a tendance à augmenter fortement la moyenne, tandis que le flou, jp2k et ff réduisent la moyenne et inclinent l'histogramme vers la droite.

L'histogramme jpeg est à nouveau fortement affecté. Les caractéristiques d'entropie spectrale également sont également indicative ou affectés fortement du type de distorsions. Comparé à l'histogramme d'entropie spatiale, l'histogramme d'entropie spectrale distingue plus clairement l'image non dégradé de celles affectées par les distorsions.

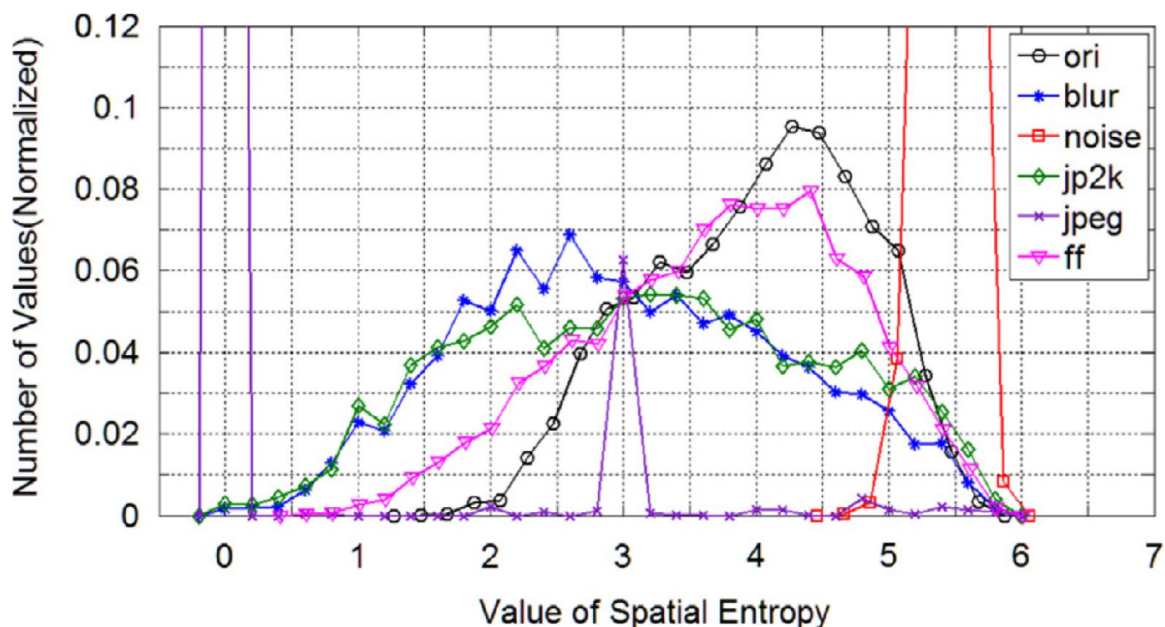
L'entropie spectrale définie ici est un descripteur précis du spectre d'énergie des images et met l'accent sur la fréquence principale et les principales orientations au sein d'un patch local. Ainsi, il est capable de distinguer le bruit et l'effet de flou plus clairement. De plus, l'entropie spectrale peut capturer la texture variations plus efficacement, auxquelles la perception humaine est très sensible.

Comme pour l'entropie spatiale, nous utilisons la moyenne et biais en tant que caractéristiques descriptive de qualité. Nous extraire ces 2 caractéristiques de chaque échelle, ce qui donne  $2 \times 3 = 6$  fonctionnalités.

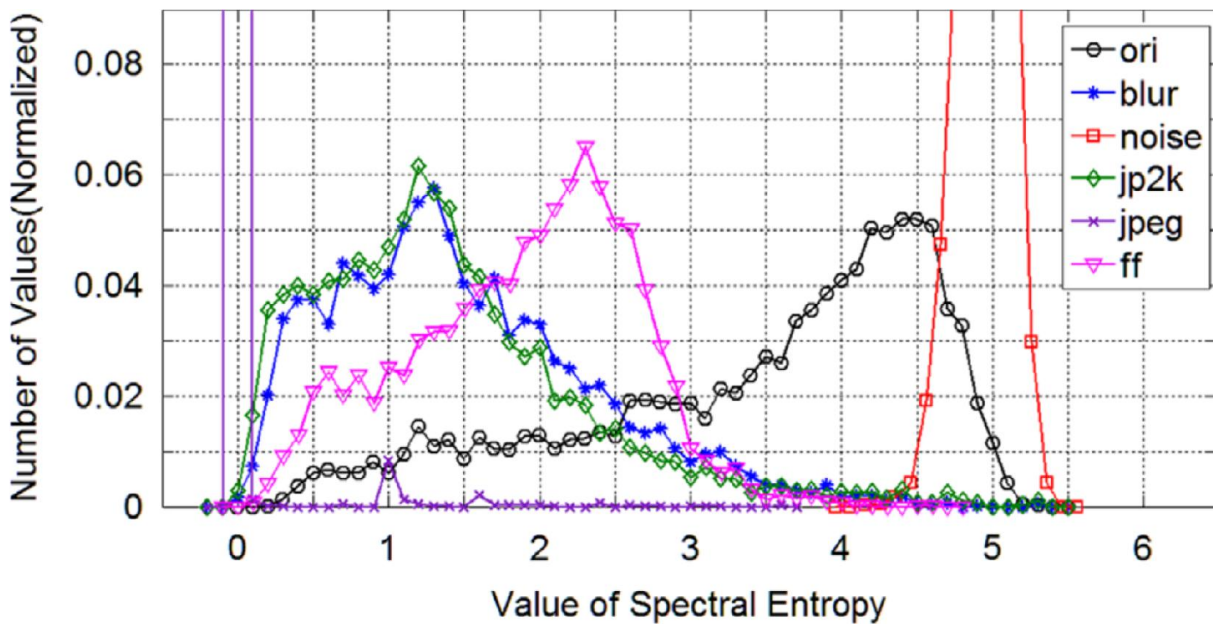
Dans les deux Fig.4.26 et 4.27 nous constatons en outre que les distributions d'entropie de «noise» et «jpeg» peuvent être clairement distinguées de celles de «ori». Parmi 5 types de distorsion, la différence entre «noise», «jpeg» et «ori» est la plus significative.

La raison en est que le «bruit» introduira beaucoup de signal haute fréquence dans le «ori», augmentant à la fois l'entropie spatiale et spectrale et «jpeg» supprimera beaucoup de signal haute fréquence du «ori», diminuant à la fois l'entropie spatiale et spectrale. De plus, «jpeg» est basé sur DCT, qui est utilisé dans notre méthode et peut être attribué à discriminer ce type de distorsion. Tandis que les distributions d'entropie de «flou» et «jp2k» ont également des formes différentes de celle de «ori» mais «flou» et «jp2k» partagent presque la même distribution d'entropie.

Cette observation est facile à comprendre que «jp2k» introduira du flou effet en supprimant de nombreuses informations à haute fréquence «ori», tandis que «décol» et «ori» ont la distribution d'entropie la plus similaire, en particulier pour la distribution d'entropie spatiale. Par conséquent, la similitude entre «décol» et «ori» se traduira par des performances relativement moins bonnes que d'autres.



**Figure 4.26** Histogrammes des valeurs d'entropie spatiale pour différents types de Distorsion. Les 6 courbes correspondent à une image non dégradée et son homologues dégradées. Remarquez comment chaque distorsion affecte les histogrammes des valeurs d'entropie spatiale.



**Figure. 4.27.** Histogrammes des valeurs d'entropie spectrale pour différents types de Distorsion. Les 6 courbes correspondent à une image non dégradée et son homologues dégradés. Les images ici sont les mêmes que celles de la Fig. 4.26. Remarquez comment chaque distorsion affecte les histogrammes des valeurs d'entropie spectrale.

Dans l'ensemble, notre modèle n'en extrait que 12 « soucieux de la qualité » caractéristiques de l'image dégradée à tester, comme indiqué dans le tableau 1. En raison de l'extraction de chaque type d'entités à partir de 3 échelles (basse, moyenne et haute), chaque groupe d'entités comprend 3 caractéristiques.

Vecteur de caractéristique	Description de caractéristique
f1-f3	Moyennes des valeurs d'entropie spatiale pour 3 échelles
f4-f6	Inclinaison des valeurs d'entropie spatiale pour 3 échelles
f7-f9	Moyennes des valeurs d'entropie spectrale pour 3 échelles
f10-f12	Inclinaison des valeurs d'entropie spectrale pour 3 échelles

**Table 1:** caractéristiques utilisées pour SSEQ (dubbed Spatial–SpectralEntropy-based Quality)

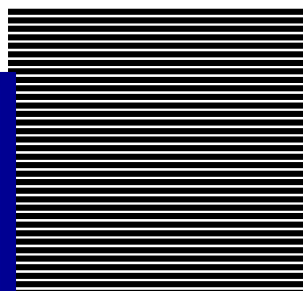
### Conclusion :

Dans ce chapitre une nouvelle métrique a été proposée. Où de multiple distorsions affectent l'image S3D , Cette nouvelle métrique est basé sur l'extraction des caractéristiques d'entropie spatiale et spectrale d'image S3D avec l'incorporation de la vision binoculaire, les puissantes machines Learning SVR ont été adoptées comme outil pour prédire ou dériver le score de la qualité.

# Chapitre

*Résultats et Discussions*

5



## Chapitre 5 : Résultats et Discussions

### 5.1 Introduction :

Dans ce chapitre on va évaluer les performances de la métrique proposée (méthode d'évaluation sans référence de la qualité de vidéo stéréoscopique basé sur l'entropie spéciale et spectrale de l'image), et pour cela des testes subjectifs et objectifs ont été mis en place, nous évaluons la mesure objective en terme de corrélation avec MOS (Mean opinion score).

L'évaluation des performances se fait en calculant les coefficients de corrélation de Pearson (CC), les coefficients corrélation des rangs de Spearman (ROCC) et l'erreur quadratique moyenne (RMSE), on va aussi évaluer les performances de notre métrique proposée en la comparant avec les métriques suivantes : (PSNR [2], SSIM [3], MS-SSIM [4], IW-SSIM [5] et VQM [6]).

### 5.2 Base de qualité de vidéo 3D Waterloo-IVC Phase I:

La base de qualité de vidéo 3D Waterloo-IVC Phase I est composée de 176 vidéos WaterlooIVC3DVideoPhaseI dégradées.

La nouvelle base de qualité de vidéo 3D Waterloo-IVC Phase I [1] est créée à partir de 4 vidéos 3D à vues multiples, à savoir Balloons, Book, Kendo et Lovebird, qui sont utilisées en commun 3D HEVC séquences de test.

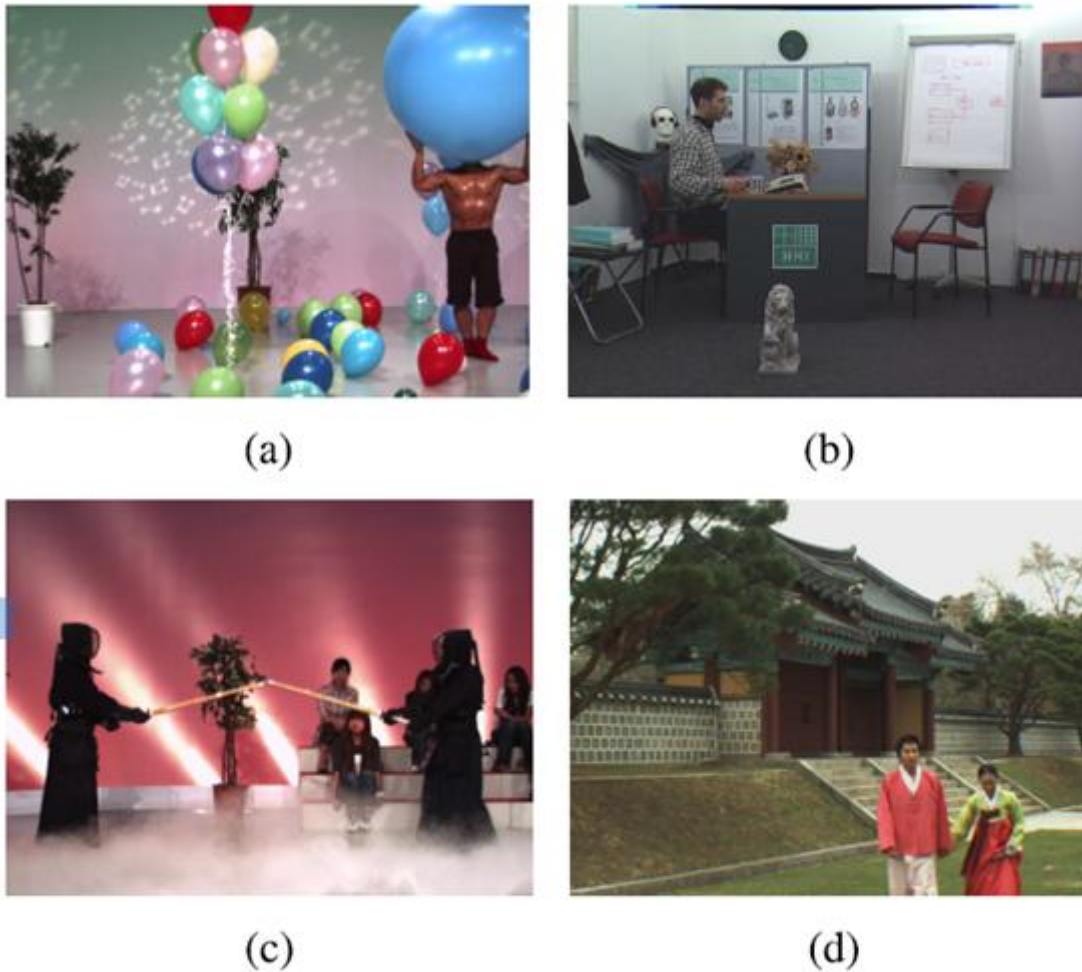
Toutes les séquences vidéo source de la Phase I ont été examinées et sélectionnées par plusieurs membres du laboratoire "Image and Vision Computing" de l'Université de Waterloo pour s'assurer qu'elles ont de bons effets 3D et ne causent pas de l'inconfort visuel. Les détails de toutes les vidéos de test sont donnés dans le tableau 1.

Toutes les vidéos sont au format YUV4: 2: 0, Des exemples pour chaque séquence de test sont illustrés à la Figure 5.1..

La base de données vidéo 3D Waterloo-IVC Phase I comprend des vidéos 3D stéréoscopiques obtenues à partir d'un codage de quantification symétrique et asymétrique du domaine de transformation suivi de différents niveaux de filtrage passe-bas. Chaque vidéo a été compressée à l'aide d'un encodeur HEVC par cinq niveaux de quantification de domaine de transformation avec  $QP = \{25, 35, 40, 45, 50\}$ . Les vidéos ont été utilisées pour générer des vidéos stéréoscopiques compressées, symétrique ou asymétrique.

	Resolution	Length	Frames/Second	Views
Book	1024×768	6s	16.67	View 6 & View 8
Balloons	1024×768	10s	30.00	View 1 & View 3
Kendo	1024×768	10s	30.00	View 1 & View 3
Lovebird	1024×768	10s	30.00	View 4 & View 6

**Tableau 5.1** La base WATERLOO-IVC 3D VIDEO PHASE I



**Figure 5.1.** Exemples d'images de vidéos. Seules les vues de droite sont affichées ici.

(a) Balloons, (b) Book, (c) Kendo, (d) Lovebird.

### 5.3 Environnement de programmation et d'implémentation :

Pour l'implémentation de la métrique proposée (méthode d'évaluation sans référence de la qualité de vidéo stéréoscopique basé sur l'entropie spéciale et spectrale de l'image) on a choisi le matlab R2014a comme un environnement d'implémentation et de programmation.

**MATLAB** (« matrix laboratory ») est un langage de script émulé par un environnement de développement du même nom ; il est utilisé à des fins de calcul numérique. Développé par la société The MathWorks, MATLAB permet de manipuler des matrices, d'afficher des courbes et des données, de mettre en œuvre des algorithmes, de créer des interfaces utilisateurs, et peut s'interfacer avec d'autres langages comme le C, C++, Java, et Fortran.

Concernant le SVR, la bibliothèque SVM LIBSVM est utilisée pour l'implémentation de la métrique proposée. Le SVR a besoins d'un apprentissage qui est nécessaire pour la phase de test. Afin d'assurer un bon apprentissage, chaque base est décomposée aléatoirement en deux sous-groupes. Où 80% des images(vidéos) est utilisées pour l'apprentissage et les 20% qui restent pour les tests. Cela est

fait pour garantir une meilleure représentation des différentes dégradations. La procédure apprentissage-test est répétée 1000 fois et la moyenne de ces 1000 itérations est donnée.

## 5.4 Critère d'évaluation de qualité d'image (vidéo):

### 5.4.1 Indice d'ajustement de qualité :

La performance d'un modèle de qualité d'image (IQA) est généralement évaluée à partir de trois aspects concernant son pouvoir de prédiction précision de la prédiction, la prévision monotonie et l'uniformité de prédiction. Le calcul de ces indices nécessite une procédure de régression pour réduire la non-linéarité des scores prédits. Nous désignons par  $Q$ ,  $Q_p$  et  $S$ : les valeurs originales d'IQA, les scores d'IQA après la régression et les notes subjectives, respectivement. La fonction de régression logistique est utilisée pour la régression non linéaire :

$$Q_p = \beta_1 \left( \frac{1}{2} - \frac{1}{\exp(\beta_2(Q - \beta_3))} \right) + \beta_4 Q + \beta_5 \quad (5.1)$$

Où  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$  et  $\beta_5$  sont les paramètres du modèle de régression

Après la régression, deux indices de correspondance peuvent être calculées pour l'évaluation des performances. La première est le coefficient de corrélation linéaire de Pearson (PCC) entre  $Q_p$  et  $S$ , qui est d'évaluer la prédiction précision La deuxième indice est l'erreur quadratique moyenne (RMSE) entre  $Q_p$  et  $S$ , qui est d'évaluer la prédiction de cohérence.

### 5.4.2 Coefficient de corrélation :

En probabilités et en statistiques, étudier la corrélation entre deux ou plusieurs variables aléatoires ou statistiques numériques, c'est étudier l'intensité de la liaison qui peut exister entre ces variables :

$$PCC(Q_p, S) = \frac{\bar{Q}_p^T \bar{S}}{\sqrt{\bar{Q}_p^T \bar{Q}_p \bar{S}^T \bar{S}}} \quad (5.2)$$

Où  $\bar{Q}_p$  et  $\bar{S}$  sont les vecteurs de moyennes, respectivement, et l'indice "T" désigne la transposition.

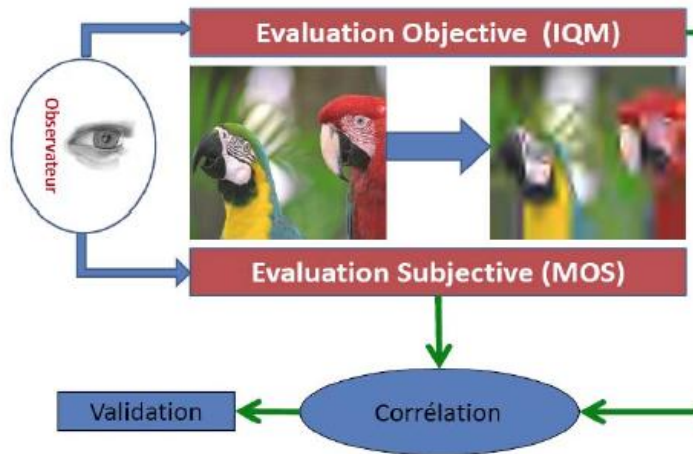


Figure 5.2 : Evaluation de la qualité.

### 5.4.3 Coefficient de corrélation des rangs de Spearman (ROCC) :

Cette mesure de corrélation présume seulement une relation monotone entre les deux quantités. Il est défini par :

$$ROCC = 1 - \frac{6 \sum (S - Q_p)^2}{n(n^2 - 1)} \quad (5.3)$$

$n$  dénote le nombre d'échantillons.

### 5.4.4 Erreur quadratique moyenne (RMSE) :

RMSE est une mesure fréquemment utilisée pour les différences entre les valeurs (l'échantillon et les valeurs de la population) prédites par un modèle ou un estimateur et les valeurs réellement observées. Fondamentalement, le RMSE représente l'écart type d'échantillon de la différence entre les valeurs prédites et les valeurs observées.

Ces différences individuelles sont appelés résidus lorsque les calculs sont effectués sur l'échantillon de données qui a été utilisée pour l'estimation, et sont appelées erreurs de prédiction lorsque calculée hors-échantillon. Le RMSE sert à agréger les grandeurs des erreurs dans les prévisions pour diverses reprises en une seule mesure de la puissance prédictive. RMSE est une bonne mesure de la précision, mais seulement de comparer les

Erreurs de prévision de modèles différents pour une variable particulière et non entre les variables :

$$RMSE(Q_p, S) = \sqrt{(Q_p - S)^T (Q_p - S) / n} \quad (5.4)$$

### 5.4.5 Significations des Coefficients :

Ces mesures (indices d'évaluation) évaluent un modèle d'aptitude objectif fournir des prévisions toujours parfait pour tous les types des images et ne manquent pas très pour un sous-ensemble d'images, c'est à dire : la cohérence dans la prédiction.

ROCC évalue la monotonie de prédiction. CC et RMSE évaluer la précision de la prédiction.

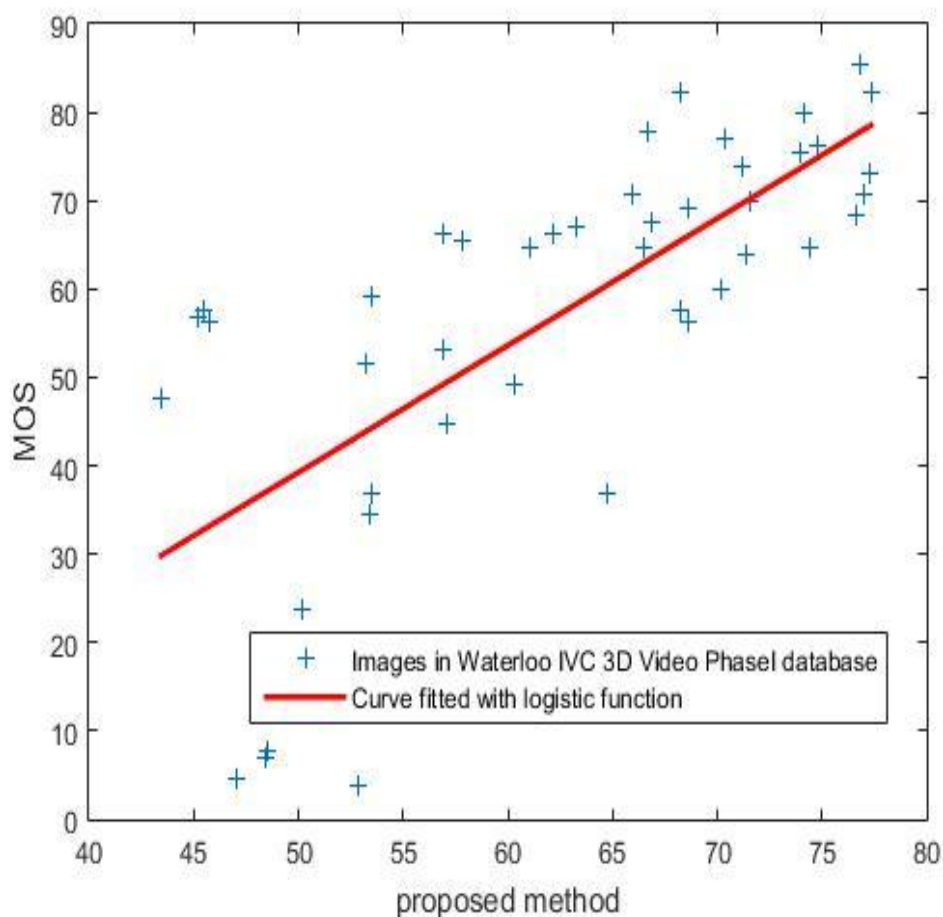
- ROCC et CC signifient une meilleure précision si leurs valeurs adjacent à 1 ou -1.
- Les plus petites valeurs de RMSE signifient que la performance est meilleure.

### 5.5 Résultat de méthode proposée :

La méthode de l'évaluation de la qualité d'image basé sur l'entropie spéciale et spectrale d'image(vidéo) a été testée sur la base vidéo Waterloo IVC 3D Video Phase I.

#### 5.5.1 Performance et comparaison :

La méthode proposée est comparée à différentes mesures (PSNR [2], SSIM[3], MS-SSIM [4], IW-SSIM [5] et VQM [6]). Les diagrammes suivants affiche les résultats obtenus en termes de corrélation avec les notes subjectives pour les différentes mesures. (Figure 5.3)



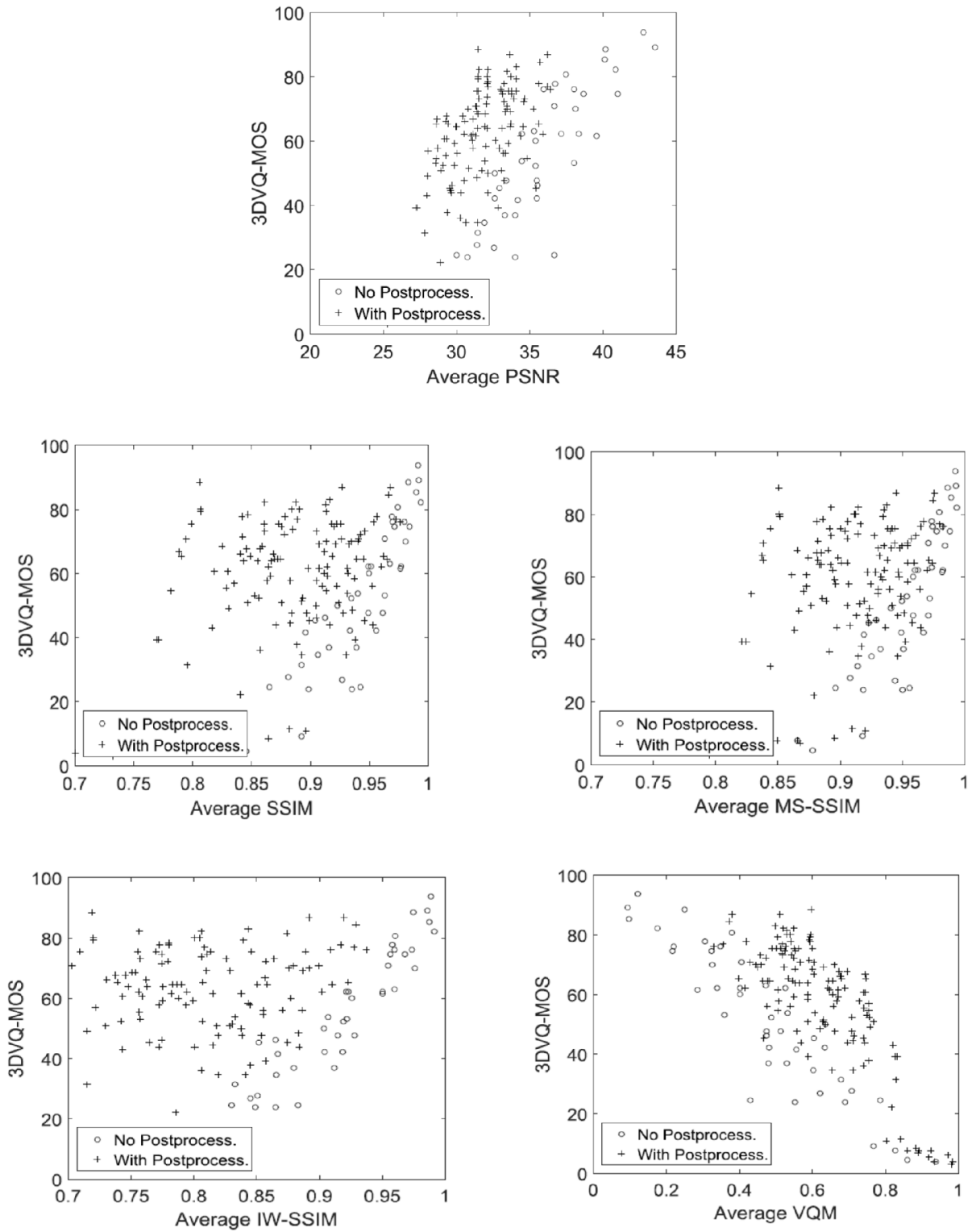


Figure 5.3: Plot des scores objectifs pour les images avec les MOS correspondant

Le tableau 5.2 montre les valeurs de coefficients de corrélation de Pearson(CC), les coefficients corrélation des rangs de Spearman (ROCC) et l'erreur quadratique moyenne (RMSE) pour la méthode proposée et les différentes mesures(PSNR [2], SSIM[3], MS-SSIM [4], IW-SSIM [5] et VQM [6]).

Méthode	PCC	ROCC	RMSE
PSNR(moyenne)	0.7085	0.5336	15.4507
SSIM(moyenne)	0.3964	0.2872	20.1010
MS-SSIM(moyenne)	0.4072	0.2969	19.9978
IW-SSIM(moyenne)	0.4833	0.2787	19.1683
VQM(moyenne)	0.7912	0.6321	13.3905
Méthode proposée	0.7416	0.7756	14.1079

**Tableau 5.2:** Comparaison des performances des modèles de prédiction de qualité 2d à 3d sur la base WATERLOO-IVC 3D VIDEO

On constate que la méthode proposée donne des résultats de corrélation mieux que les autres méthodes, où les corrélations les plus élevées sont obtenus par la méthode proposée par rapport aux autres méthodes.

Alors notre méthode proposée donne des résultats satisfaites en terme de corrélation.

On constate aussi que dans la méthode proposée les scores objectifs sont à peu près proche de ce lui des scores subjectifs.

La Figure 5.3 montre que les valeurs de la métrique proposée et PSNR sont plus proches aux valeurs des notes subjectives.

Les valeurs de ROCC et CC pour toutes les mesures sont entre l'intervalle  $[-1,1]$ , cela signifie une meilleure précision de prédiction.

Pour les deux mesures (méthode proposé et VQM) la performance est meilleure par rapport aux autre mesures.

En terme de l'indice PCC les méthodes (proposée, PSNR et VQM) sont proche.

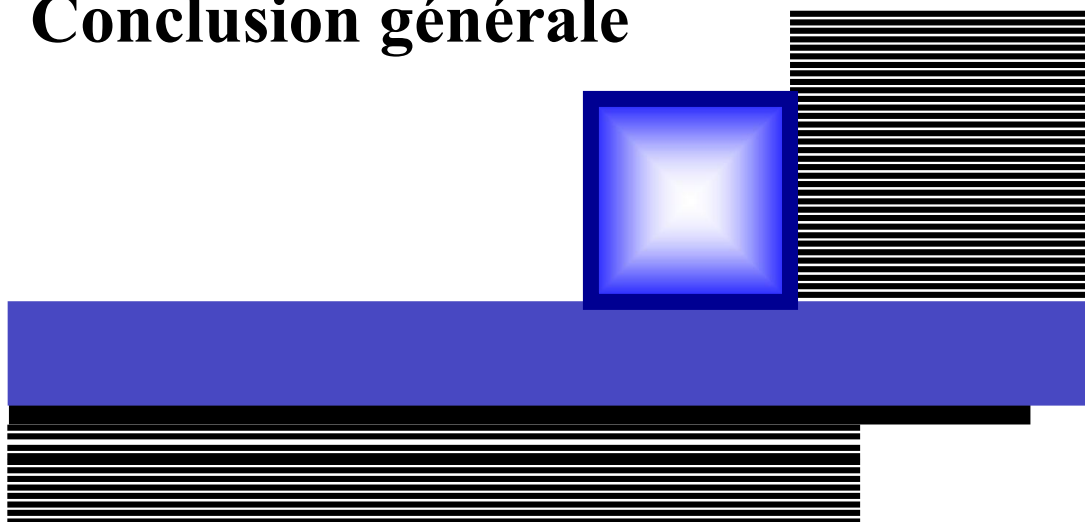
En terme de l'indice ROCC les méthodes (proposée et VQM) sont proche.

## 5.6 conclusion :

Dans ce chapitre, nous avons testé la performance de la méthode de l'évaluation sans référence de la qualité de vidéo stéréoscopique 3D basé sur les caractéristiques d'entropie spatiale et spectrale de l'image. La base de qualité de vidéo 3D Waterloo-IVC Phase I a été utilisée dans l'expérience. En plus, nous avons calculé les indices PCC, ROCC et RMSE.

Une étude comparative entre, les mesures PSNR, SSIM, MS-SSIM, IW-SSIM et VQM et la méthode proposée a été menée. Cette comparaison nous a conduit à tirer la conclusion que la méthode proposée donne des résultats satisfaisants et prometteurs par rapport aux autres méthodes notamment en terme de corrélation entre les valeurs subjectif MOS et les valeurs objectif..

# Conclusion générale



## **Conclusion générale :**

Le travail entre nos mains présente une étude dans laquelle le sujet de l'évaluation de qualité d'image est traité. Vu l'importance de l'évaluation pour la recherche ou bien pour les applications temps réel, nous nous sommes intéressés l'évaluation sans référence sur des vidéos stéréoscopiques, notre méthode sans référence est utilisée pour mesurer les vidéos stéréoscopiques dégradées par différentes distorsions, elle est basée sur les caractéristiques d'entropie spatiale et spectrale d'image, Cette fois-ci, les caractéristiques du système visuel humain sont introduites. l'image stéréoscopique perçu est le résultat de la combinaison des deux images gauche et droite par un model de combinaison.

Les machines à apprentissage tel que les machines à vecteur de support SVR représente un outil puissant, capable de produire à partir des scores objectifs et subjectifs un score unique qui décrit la qualité de l'image.

Les résultats obtenus à travers l'application du model sont témoins de l'efficacité et la performance de notre modèle.

La méthode proposée est appliquée sur des base de videos stéréoscopiques 3D Waterloo-IVC Phase I, et elle utilise la machine Learning SVM pour prédire de scores de qualité, d'autre métrique peut travailler sur d'autre base de vidéo différente que la base 3D Waterloo-IVC Phase I et utilise aussi les réseaux de neurone pour produire des scores de qualité d'image.

## BIBLIOGRAPHIE

- [1] J. Wang, S. Wang, and Z. Wang, "Quality prediction of asymmetrically compressed stereoscopic videos," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, QC, Canada, Sep. 2015, pp. 1–5.
- [2] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Process., Image Commun.*, vol. 28, no. 8, pp. 870–883, Dec. 2013.
- [3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [4] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2003, pp. 1398–1402.
- [5] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
- [6] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.
- [7] THÈSE de doctorat : « Évaluation de la qualité perceptuelle des signaux multimédias » par : DAKKAR BORHEN EDDINE 2017.
- [8] THÈSE de doctorat : «Evaluation de la qualité d'image» par : AHMED SEGHIR Zianou 2012.
- [9] Rapport Bibliographique - Master Recherche "Architectures Logicielles Distribuées" «Etat de l'art des critères objectifs de qualité d'images et de vidéos» 2006.
- [10] «Une nouvelle métrique avec référence complète d'évaluation de la qualité d'images couleur» par Ezzeddine Zagrouba 2012.
- [11] <http://thesis.univ-biskra.dz/2271/6/Chapitre%2003.pdf>.
- [12] [https://www.researchgate.net/publication/269336863\\_Exploitation\\_des\\_caracteristiques\\_du\\_systeme\\_visuel\\_humain\\_dans\\_les\\_metriques\\_de\\_qualite](https://www.researchgate.net/publication/269336863_Exploitation_des_caracteristiques_du_systeme_visuel_humain_dans_les_metriques_de_qualite).
- [13] <https://tel.archives-ouvertes.fr/tel-00560808/document>
- [14] <https://hal.inria.fr/hal-01286650/document>
- [15] <http://deptinfo.unice.fr/~ol/L1MI-UA/memo6.pdf>
- [16] [https://www.univ-usto.dz/images/coursenligne/these\\_cherifi\\_daikha.pdf](https://www.univ-usto.dz/images/coursenligne/these_cherifi_daikha.pdf)
- [17] [http://www.academia.edu/10339118/Evaluation\\_de\\_la\\_qualit%C3%A9\\_des\\_images](http://www.academia.edu/10339118/Evaluation_de_la_qualit%C3%A9_des_images)
- [18] [https://www.image-enrelief.org/stereo/images/documents/Breve\\_histoire\\_stereoscopie.pdf](https://www.image-enrelief.org/stereo/images/documents/Breve_histoire_stereoscopie.pdf)
- [19] <http://www.stereoscopie.eu/expo1/Conf1.pdf>
- [20] <https://fr.wikipedia.org/wiki/St%C3%A9r%C3%A9oscopie>
- [21] [https://fr.wikipedia.org/wiki/Image\\_st%C3%A9r%C3%A9oscopique](https://fr.wikipedia.org/wiki/Image_st%C3%A9r%C3%A9oscopique)
- [22] <https://www.memoireonline.com/04/08/1033/stereoscopie-art-de-la-vision-en-relief.html>
- [23] <http://maths.cnam.fr/Stagiaires/CS/ArnaudLitrico/IMG/pdf/appariement.pdf>

- [24] <https://tel.archives-ouvertes.fr/tel-00011437/document>
- [25] <https://www.it-swarm.dev/fr/computer-vision/difference-entre-la-carte-de-disparite-et-limage-de-disparite-dans-la-correspondance-stereo/1041258396/>
- [26] [https://fr.qwe.wiki/wiki/Binocular\\_disparity](https://fr.qwe.wiki/wiki/Binocular_disparity)
- [27] <https://core.ac.uk/download/pdf/35402383.pdf>
- [28] <http://thesis.univ-biskra.dz/2097/>
- [29] <http://hal.inria.fr/docs/00/11/03/44/PDF/LauerBlochJIME06.pdf>
- [30] [https://fr.wikipedia.org/wiki/Machine\\_%C3%A0\\_vecteurs\\_de\\_support](https://fr.wikipedia.org/wiki/Machine_%C3%A0_vecteurs_de_support)
- [31] [https://fr.wikipedia.org/wiki/Rivalit%C3%A9\\_binoculaire](https://fr.wikipedia.org/wiki/Rivalit%C3%A9_binoculaire)
- [32] <https://en.wikipedia.org/wiki/Stereopsis>
- [34] [https://fr.wikipedia.org/wiki/Vision\\_binoculaire#:~:text=La%20vision%20binoculaire%20est%20un,et%20oculus%20pour%20%C2%AB%20yeux%20%C2%BB](https://fr.wikipedia.org/wiki/Vision_binoculaire#:~:text=La%20vision%20binoculaire%20est%20un,et%20oculus%20pour%20%C2%AB%20yeux%20%C2%BB)
- [35] article1 : « Binocular spatial activity and reverse saliency driven no-reference stereopair quality assessment »
- [36] article2 : « No-reference image quality assessment based on spatial and spectral entropies »
- [37] THÈSE de doctorat : « Une méthode pour l'évaluation de la qualité des images 3D stéréoscopiques »  
Mlle Raluca VLAD en 2006